

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

QM protein - protein search, using sw model

Run on: June 23, 2005, 08:12:43 ; Search time 42.1481 Seconds
(without alignments)
3578.727 Million cell updates/sec

Title: US-10-697-828-8

Perfect score: 2051

Sequence: 1 MWLPRFSKTVTVLLAQTT.....LTDLVLPRGDFHFSWASPD 390

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq_16Dec04:*

- 1: geneseqp1980s:*
- 2: geneseqp1990s:*
- 3: geneseqp2000s:*
- 4: geneseqp2001s:*
- 5: geneseqp2002s:*
- 6: geneseqp2003as:*
- 7: geneseqp2003bs:*
- 8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|--------------------|
| 1 | 2051 | 100.0 | 390 | 4 | AAY72639 Human gly |
| 2 | 2051 | 100.0 | 390 | 5 | Abb81556 Human int |
| 3 | 1865.5 | 91.0 | 418 | 3 | Aab41947 Human ORF |
| 4 | 1758.5 | 85.7 | 395 | 5 | Abb81555 Consensus |
| 5 | 1729.5 | 84.3 | 395 | 4 | AAY72640 Human gly |
| 6 | 1729.5 | 84.3 | 395 | 5 | Abb81554 Human cor |
| 7 | 1729.5 | 84.3 | 395 | 5 | Aae15438 Human dru |
| 8 | 1729.5 | 84.3 | 395 | 7 | Adi21086 Novel hum |
| 9 | 1729.5 | 84.3 | 395 | 8 | Adl61235 Human tyr |
| 10 | 1536.5 | 74.9 | 395 | 4 | AAY72638 Mouse gly |
| 11 | 1536.5 | 74.9 | 395 | 5 | AAU11275 Murine in |
| 12 | 1523 | 74.3 | 418 | 5 | Abb81557 Mouse int |
| 13 | 1513.5 | 73.8 | 394 | 7 | ADJ70405 Human hea |
| 14 | 1013 | 49.4 | 386 | 4 | AAM93309 Human pol |
| 15 | 1013 | 49.4 | 386 | 8 | ADL30784 Human pro |
| 16 | 1008 | 49.1 | 386 | 2 | AAV39918 Human gly |
| 17 | 1003.5 | 48.9 | 380 | 5 | AAU11274 Human L-s |
| 18 | 970 | 47.3 | 386 | 3 | AAV79219 Human tra |
| 19 | 958.5 | 46.7 | 388 | 5 | AAV39919 Mouse gly |
| 20 | 808 | 39.4 | 169 | 5 | Abb81559 Human int |
| 21 | 755 | 36.8 | 169 | 5 | Abb81558 Human cor |
| 22 | 665 | 32.4 | 483 | 2 | AAV31656 Mouse N-a |
| 23 | 659.5 | 32.2 | 530 | 4 | AAB95367 Human pro |
| 24 | 659.5 | 32.2 | 530 | 8 | ADQ18590 Human sof |
| 25 | 656.5 | 32.0 | 484 | 2 | AAV31657 Human N-a |

ALIGNMENTS

RESULT 1

AAV72639
ID AAY72639 standard; protein; 390 AA.

AC AAY72639;

XX 02-MAY-2001 (first entry)

DE Human glycosyl sulfotransferase-4alpha (GST-4alpha).

XX Human; glycosyl sulfotransferase-4alpha; GST-4alpha; immunosuppressive; therapy; selectin binding inhibitor; gene therapy; inflammation;
KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
KW polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenalitis;
KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
KW asthma; hypersensitivity; rheumatic fever; tissue rejection;
KW chromosome 16q23.1.

XX Homo sapiens.

OS WO200106015-A1.

XX 25-JAN-2001.

PF 19-JUL-2000; 2000WO-US019741.

XX 20-JUL-1999; 99US-0144694P.

PR 13-JUN-2000; 2000US-00593828.

XX (REGC) UNIV CALIFORNIA.

XX Rosen SD, Lee JK, Hemmerich S;

XX WPI; 2001-138471/14.

XX N-PSDB; AAD02697, AAD02698, AAD02699.

XX New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for

diagnostic and therapeutic agent screening applications.

XX Claim 3; Fig 1; 128pp; English.

XX The present sequence is human glycosyl sulfotransferase-4alpha (GST-4

alpha). GST-4 gene is found on chromosome 16q23.1. GST is a type 2

membrane protein useful for inhibiting a binding event between a selectin

and a selectin ligand, which comprises contacting the selectin with a non

CC -sulphated selectin ligand, GST and a small molecular agent that inhibits
 CC the sulphation activity of GST. GST is also useful in inhibiting a
 CC selectin mediated binding event. GST is useful in gene therapy to treat
 CC disorders such as acute or chronic inflammation, systemic lupus
 CC erythematosus (SLE), rheumatoid arthritis, polyarteritis nodosa,
 CC polymyositis, dermatomyositis, systemic sclerosis, diabetes,
 CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
 CC disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious
 CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
 CC dermatitis, myocarditis, regional enteritis, adult respiratory distress
 CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
 CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
 CC during transplantation. Note: The present sequence is also shown in
 CC sequence listing (page no: 56) but lacks four nucleotides at its 3' end
 XX
 XX

XX Sequence 390 AA;

Query Match 100.0%; Score 2051; DB 4; Length 390;
 Best Local Similarity 100.0%; Pred. No. 1.2e-215;
 Matches 390; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MWLPFSSKTVTVLLLAQTTCCLLFIISRPSPSPAGGEDRVHVLVLSWRSGSFLGQL 60
 Db 1 MWLPFSSKTVTVLLLAQTTCCLLFIISRPSPSPAGGEDRVHVLVLSWRSGSFLGQL 60
 Qy 61 FSOHPDVFYLMPEPAWHVMTTLSQSAATLHMAVRDLMSIFLCMDVFDAYMPSQSNLSA 120
 Db 61 FSOHPDVFYLMPEPAWHVMTTLSQSAATLHMAVRDLMSIFLCMDVFDAYMPSQSNLSA 120
 Qy 121 FFWNATSRALCSPACSAFPRGTISKQDVCKTLCTRPFSLAREACRSYSHVVLKEVRFF 180
 Db 121 FFWNATSRALCSPACSAFPRGTISKQDVCKTLCTRPFSLAREACRSYSHVVLKEVRFF 180
 Qy 181 NLQVLYPLSDPALNLRIIVHLVRDPAVLRSAAGPILARDNGIVLGTNGKWEADPHL 240
 Db 181 NLQVLYPLSDPALNLRIIVHLVRDPAVLRSAAGPILARDNGIVLGTNGKWEADPHL 240
 Qy 241 RLIREVCRSHVRVIAEAATLKPPPLRGYRLVRFPEDLAREPLAEIRALYFTGLTLPOL 300
 Db 241 RLIREVCRSHVRVIAEAATLKPPPLRGYRLVRFPEDLAREPLAEIRALYFTGLTLPOL 300
 Qy 301 EAWIHNITHGSGIGKPIEAFHTSSRNARNVSQAWRHALPFTKILRQVEVCAGALQLGYR 360
 Db 301 EAWIHNITHGSGIGKPIEAFHTSSRNARNVSQAWRHALPFTKILRQVEVCAGALQLGYR 360
 Qy 361 PVYSADQQRDLTDLVLPGRGPDHFSWASPD 390
 Db 361 PVYSADQQRDLTDLVLPGRGPDHFSWASPD 390

RESULT 2

ABB81556 ID ABB81556 standard; protein; 390 AA.

AC ABB81556;

XX 05-SEP-2002 (first entry)

XX Human intestinal N-acetylglucosamine-6-sulfotransferase SEQ ID NO:4.
 XX Human; N-acetylglucosamine-6-sulfotransferase; enzyme; GlcNAc6ST;
 XX corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
 XX ophthalmological.

OS Homo sapiens.

XX US2002061562-A1.

XX 23-MAY-2002.

XX 09-AUG-2001; 2001US-00927602.

XX 11-AUG-2000; 2000US-00638211.

PR 11-AUG-2000; 2000US-0325773P.

XX (FUKU/) FUKUDA M N.

PA (AKAW/) AKAWA T O.

XX Fukuda MN, Akama TO;

XX WPI; 2002-507643/54.

XX New nucleic acid encoding corneal N-acetylglucosamine-6-sulfotransferase,
 PT useful for treatment, monitoring and diagnosis of macular corneal
 PT dystrophy.

XX Example 5; Fig 2A-B; 69pp; English.

XX The present invention describes human corneal N-acetylglucosamine-6-
 CC sulfotransferase (I), which is able to catalyse sulfation of keratan
 CC sulfate (KS). Also described is a method for monitoring the effect of
 CC treatments for macular corneal dystrophy (MCD), and detecting
 CC susceptibility to MCD. (I) is located to chromosome 16q22, and has
 CC ophthalmological activity. (I) can be used to treat or prevent macular
 CC corneal dystrophy types I or II. (I) makes possible treatment of MCD
 CC without requiring keratoplasty or keratectomy. The present sequence
 CC represents human intestinal N-acetylglucosamine-6- sulfotransferase,
 CC which is given in comparison with (I) in the exemplification of the
 CC present invention

XX Sequence 390 AA;

Query Match 100.0%; Score 2051; DB 5; Length 390;
 Best Local Similarity 100.0%; Pred. No. 1.2e-215;
 Matches 390; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MWLPFSSKTVTVLLLAQTTCCLLFIISRPSPSPAGGEDRVHVLVLSWRSGSFLGQL 60
 Db 1 MWLPFSSKTVTVLLLAQTTCCLLFIISRPSPSPAGGEDRVHVLVLSWRSGSFLGQL 60
 Qy 61 FSOHPDVFYLMPEPAWHVMTTLSQSAATLHMAVRDLMSIFLCMDVFDAYMPSQSNLSA 120
 Db 61 FSOHPDVFYLMPEPAWHVMTTLSQSAATLHMAVRDLMSIFLCMDVFDAYMPSQSNLSA 120
 Qy 121 FFWNATSRALCSPACSAFPRGTISKQDVCKTLCTRPFSLAREACRSYSHVVLKEVRFF 180
 Db 121 FFWNATSRALCSPACSAFPRGTISKQDVCKTLCTRPFSLAREACRSYSHVVLKEVRFF 180
 Qy 181 NLQVLYPLSDPALNLRIIVHLVRDPAVLRSAAGPILARDNGIVLGTNGKWEADPHL 240
 Db 181 NLQVLYPLSDPALNLRIIVHLVRDPAVLRSAAGPILARDNGIVLGTNGKWEADPHL 240
 Qy 241 RLIREVCRSHVRVIAEAATLKPPPLRGYRLVRFPEDLAREPLAEIRALYFTGLTLPOL 300
 Db 241 RLIREVCRSHVRVIAEAATLKPPPLRGYRLVRFPEDLAREPLAEIRALYFTGLTLPOL 300
 Qy 301 EAWIHNITHGSGIGKPIEAFHTSSRNARNVSQAWRHALPFTKILRQVEVCAGALQLGYR 360
 Db 301 EAWIHNITHGSGIGKPIEAFHTSSRNARNVSQAWRHALPFTKILRQVEVCAGALQLGYR 360
 Qy 361 PVYSADQQRDLTDLVLPGRGPDHFSWASPD 390
 Db 361 PVYSADQQRDLTDLVLPGRGPDHFSWASPD 390

RESULT 3

AAB41947

ID AAB41947 standard; protein; 418 AA.

XX AAB41947;

XX 08-FEB-2001 (first entry)

XX Human ORFX ORF1711 polypeptide sequence SEQ ID NO:3422.

XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;

KW vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW immunosuppressant; osteoporotic; antidiabetic; immunosuppressant; cardiac;
KW immunosuppressant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antineoplastic; antithyroid;
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.
OS Homo sapiens.
XX WO2000058473-A2.
XX 05-OCT-2000.
XX 31-MAR-2000; 2000WO-US008621.
XX 31-MAR-1999; 99US-0127607P.
XX 02-APR-1999; 99US-0127636P.
XX 05-APR-1999; 99US-0127728P.
XX 30-MAR-2000; 2000US-00540763.
XX (CURA-) CURAGEN CORP.
XX Shimkets RA, Leach M;
XX WPI; 2000-602362/57.
XX N-PSDB; AAC76156.
XX Novel nucleic acids and peptides derived from open reading frame X,
XX useful for treating e.g. cancers, proliferative disorders,
XX neurodegenerative disorders and cardiovascular disease.
XX Claim 11; Page 2599-2600; 5507pp; English.
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
XX which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX sequences have activities such as: cytostatic; hepatotropic; vulnery;
XX antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic;
XX anticonvulsant; antidiabetic; immunosuppressant; immunostimulant;
XX cardiac; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
XX dermatological; immunosuppressive; antiinflammatory; antibacterial;
XX antiviral; antifungal; antineoplastic; antithyroid; and antianemic. The
XX sequences can be used for determining the presence of or predisposition
XX to or preventing or treating pathological conditions associated with an
XX ORFX-associated disorder. The nucleic acids can be used to express ORFX
XX proteins in gene therapy vectors. The proteins and nucleic acids may be
XX used to treat cancers, proliferative disorders, neurodegenerative
XX disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
XX diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
XX storage, systemic lupus erythematosus, severe combined immunodeficiency
XX (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
XX disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
XX cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
XX enhance coagulation; to inhibit thrombosis; and as a contraceptive
SQ Sequence 418 AA;
Query Match 91.0%; Score 1865.5; DB 3; Length 418;
Beat Local Similarity 91.0%; Pred. No. 3.1e-195;
Matches 356; Conservative 9; Mismatches 25; Indels 1; Gaps 1;
QY 1 MWLPRFSKTVVLLAQTCLLLFIISRPSPSPAGGEDRVHVLVLSWSRSGSFLGQL 60
DB 28 MWLPRFSKTVVLLAQTCLLLFIISRPSPSPAGGEDRVHVLVLSWSRSGSFLGQL 87
QY 61 FSOHPDVYLMPEAMHWVWTTLSQSSAATLHMAVRDLMSIFLCMDVDFDAYMEGPPRRQS 119

Db 88 FSOHPDVYLMPEAMHWVWTTLSQSSAATLHMAVRDLMSIFLCMDVDFDAYMEGPPRRQS 147
QY 120 AFFNWTASRALCPACSPAPRGITISKQVCKTLCTRPQFSLAREACRSYSHVVLKEVRF 179
Db 148 SLQWENSRLCAPACDIIPQDEIIPRAHRLCSQPFVVEKACRSYSHVVLKEVRF 207
QY 180 FNLQVLYPLSDPALNURIVHLVRDPAVLRSEAAAGPILARDNGIVLGTNGKWEADPH 239
Db 208 FNLQSLYPLKDPKPSNLHLVHLVRDPAVLRSEAAAGPILARDNGIVLGTNGKWEADPH 267
QY 240 LRLIREVCRSHVRIAEATLKPPLGRYRLVRFEDLAREPLAEIRALYAFGLTLTPQ 299
Db 268 LRLIREVCRSHVRIAEATLKPPLGRYRLVRFEDLAREPLAEIRALYAFGLTLTPQ 327
QY 300 LEAWIHNTHGSGIGKPIEAFHTSSRNARNVSOAWRHALPFTKILRVQEVCAQALQLLGY 359
Db 328 LEAWIHNTHGSGIGKPIEAFHTSSRNARNVSOAWRHALPFTKILRVQEVCAQALQLLGY 387
QY 360 RPVYSADQQRDLTLDLVLRGPDHFWASPD 390
Db 388 RPVYSADQQRDLTLDLVLRGPDHFWASPD 418
RESULT 4
AAB81555
ID ABB81555 standard; protein; 395 AA.
XX ABB81555;
XX 05-SEP-2002 (first entry)
XX Consensus N-acetylglucosamine-6-sulfotransferase SEQ ID NO:3.
XX Human: N-acetylglucosamine-6-sulfotransferase; enzyme; GlcNAc6ST;
XX corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
XX ophthalmological.
XX Homo sapiens.
XX Mus musculus.
XX Synthetic.
XX Key Location/Qualifiers
FH Misc-difference 10 /label= Ala, Thr, Val
FT Misc-difference 13 /label= Ala, Val, Ser
FT Misc-difference 20 /label= Phe, Cys, Gly
FT Misc-difference 39 /label= Ala, Asp, Glu
FT Misc-difference 96 /label= Val, Met, Ile
FT Misc-difference 142 /label= Ala, Thr, Asn
FT Misc-difference 147 /label= Ala, Asp, Glu
FT Misc-difference 159 /label= Thr, Ser, Gly
FT Misc-difference 238 /label= Gly, His, Arg
FT Misc-difference 294 /label= Ser, Thr, Gly
FT Misc-difference 371 /label= Ala, Thr, Ser
FT Misc-difference 380 /label= Leu, Pro, Met
FT Misc-difference 382 /label= Gly, His, Ser
FT Misc-difference 384 /label= Thr, Ser, Lys
FT Misc-difference 390 /label= Ala, Glu
FT Misc-difference 391

FT Misc-difference 392 /label= Ser, Lys
 FT Misc-difference 394 /label= His, Gln
 FT Misc-difference 394 /label= Arg, Glu
 FT Misc-difference 395 /label= Asn, Ser
 XX
 PN US2002061562-A1.
 XX
 XX 23-MAY-2002.
 XX
 XX 09-AUG-2001; 2001US-00927602.
 XX
 XX 11-AUG-2000; 2000US-00638211.
 PR 11-AUG-2000; 2000US-0325773P.
 XX
 XX (FUKU/) FUKUDA M N.
 PA (AKAM/) AKAMA T O.
 XX
 XX Fukuda MN, Akama TO;
 PI
 XX WPI; 2002-507643/54.
 DR
 XX
 XX New nucleic acid encoding corneal N-acetylglucosamine-6-sulfotransferase,
 PT useful for treatment, monitoring and diagnosis of macular corneal
 PT dystrophy.
 XX
 XX Example 5; Fig 2A-B; 69pp; English.
 XX
 XX The present invention describes human corneal N-acetylglucosamine-6-
 CC sulfotransferase (I), which is able to catalyse sulfation of keratan
 CC sulfate (KS). Also described is a method for monitoring the effect of
 CC treatments for macular corneal dystrophy (MCD), and detecting
 CC susceptibility to MCD. (I) is located to chromosome 16q22, and has
 CC ophthalmological activity. (I) can be used to treat or prevent macular
 CC corneal dystrophy types I or II. (I) makes possible treatment of MCD
 CC without requiring keratoplasty or keratotomy. The present sequence
 CC represents a consensus N-acetylglucosamine-6-sulfotransferase which is
 CC given in the exemplification of the present invention
 XX
 XX Sequence 395 AA;
 SQ
 Query Match 85.7%; Score 1758.5; DB 5; Length 395;
 Best Local Similarity 87.1%; Pred. No. 1.6e-183;
 Matches 338; Conservative 10; Mismatches 39; Indels 1; Gaps 1;
 QY 1 MWLPFRSSKTVTVLLLAQTCLLLFIISRPSPAGGEDRVHVLVLSWRSGSSFLGQL 60
 DB 1 MWLPFRSSTVTVLLLAQTCLLLFIISRPSPAGGEDRVHVLVLSWRSGSSFLGQL 59
 QY 61 FSQHPDVFYLMPEAHVWTTLSQGSATLHMAVRDLMSIFLCMDVDFDYMPPQSRNLSA 120
 DB 60 FSQHPDVFYLMPEAHVWTTLSQGSATLHMAVRDLMSIFLCMDVDFDYMPPQSRNLSA 119
 QY 121 FFWNATSRALCSPACSAFPGRTISKQDVCKTCTROPFSLAREACSYSHVVLKEVRF 180
 DB 120 LFQNAVSRALCSPACSAFPGRTISKQDVCKTCTROPFSLAREACSYSHVVLKEVRF 179
 QY 181 NLQVLYPLLSDDPALNLRIHVLPDRAVLSRAAGPILARDNGIVLTGKNGVEADPHL 240
 DB 180 NLQVLYPLLSDDPALNLRIHVLPDRAVLSRAAGPILARDNGIVLTGKNGVEADPHL 239
 QY 241 RLIREVCRSHVRIAEAAATLKPFPFLRGYRLVFPEDLAREPLAEIRALYAFGLTLPQL 300
 DB 240 RVVREVCVSHVRIAEAAATLKPFPFLRGYRLVFPEDLAREPLAEIRALYAFGLTLPQL 299
 QY 301 EAMHNITHGSGIKPTEAFHTSSRNANVSOQWRHALPTKILRVQEVGCAGALQLGYR 360
 DB 300 EAMHNITHGSGIKPTEAFHTSSRNANVSOQWRHALPTKILRVQEVGCAGALQLGYR 359
 QY 361 PVYSADQOORDLTLDLVLPRGPDHFHSMAS 388
 DB 361 PVYSADQOORDLTLDLVLPRGPDHFHSMAS 387

Db 360 PVYSEDEQRDLXLDLVLPRGXDXFXWAS 387
 RESULT 5
 AA72640
 ID AAY72640 standard; protein; 395 AA.
 XX
 XX AAY72640;
 XX
 DT 02-MAY-2001 (first entry)
 XX
 DE Human glycosyl sulfotransferase-4beta (GST-4beta).
 XX
 KW Human; glycosyl sulfotransferase-4beta; GST-4beta; immunosuppressive;
 KW therapy; selectin binding inhibitor; gene therapy; inflammation;
 KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
 KW polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
 KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adenitis;
 KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
 KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
 KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
 KW asthma; hypersensitivity; rheumatic fever; tissue rejection;
 KW chromosome 16q23.1.
 XX
 OS Homo sapiens.
 XX
 PN WO200106015-A1.
 XX
 XX 25-JAN-2001.
 XX
 XX 19-JUL-2000; 2000WO-US019741.
 XX
 XX 20-JUL-1999; 99US-0144694P.
 PR 13-JUN-2000; 2000US-00593828.
 XX
 XX (REGC) UNIV CALIFORNIA.
 XX
 XX Rosen SD, Lee JK, Hemmerich S;
 PI WPI; 2001-138471/14.
 DR N-PSDB; AAD02697, AAD02700.
 XX
 XX New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
 PT diagnostic and therapeutic agent screening applications.
 XX
 PS Claim 3; Fig 4B; 128pp; English.
 XX
 CC The present sequence is human glycosyl sulfotransferase-4beta (GST-4
 CC beta). GST-4 gene is found on chromosome 16q23.1. GST is a type 2
 CC membrane protein useful for inhibiting a binding event between a selectin
 CC and a selectin ligand, which comprises contacting the selectin with a non
 CC -sulphated selectin ligand, GST and a small molecular agent that inhibits
 CC the sulphation activity of GST. GST is also useful in inhibiting a
 CC selectin mediated binding event. GST is useful in gene therapy to treat
 CC disorders such as acute or chronic inflammation, systemic lupus
 CC erythematosus (SLE), rheumatoid arthritis, polyarteritis nodosa,
 CC polymyositis, dermatomyositis, systemic sclerosis, diabetes,
 CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
 CC disease, Grave's disease, adenitis, hypoparathyroidism, pernicious
 CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
 CC dermatitis, myocarditis, regional enteritis, adult respiratory distress
 CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
 CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
 CC during transplantation
 XX
 XX Sequence 395 AA;
 SQ
 Query Match 84.3%; Score 1729.5; DB 4; Length 395;
 Best Local Similarity 85.8%; Pred. No. 2.4e-180;
 Matches 333; Conservative 17; Mismatches 37; Indels 1; Gaps 1;
 QY 1 MWLPFRSSKTVTVLLLAQTCLLLFIISRPSPAGGEDRVHVLVLSWRSGSSFLGQL 60
 DB 1 MWLPFRSSTVTVLLLAQTCLLLFIISRPSPAGGEDRVHVLVLSWRSGSSFLGQL 59

Db 1 MWLPRVSTAVTALLAQ-TFLLLFLVSRPGSPSPAGGEARVHVVLSSWRSGSSFVQL 59
QY 61 FSOHPDVYLMPEPAHWVTTLSQGSAAATLHMAVRDLMSIFLCMDVDFDAYMPOSRLSA 120
Db 60 FNQHPDVYLMPEPAHWVTTLSQGSAAATLHMAVRDLVRSVFLCMDVDFDAYLPWRRNLSD 119
QY 121 FFWNATSRALCSPSPACSAFPRGTISKQDVCKTLCTROPFSLAREACRSYSHVVLKEVRF 180
Db 120 LFQWAVSRALCSPSPACSAFPRGAISSEAVCKPLCARQSFTLAREACRSYSHVVLKEVRF 179
QY 181 NLQVLYPLSDPALNLRIHVLRDPRAVLRSPAGGEARVHVVLSSWRSGSSFVQL 240
Db 180 NLQVLYPLSDPALNLRIHVLRDPRAVLRSPAGGEARVHVVLSSWRSGSSFVQL 239
QY 241 RLIREVCRSHVRIAEATLKPPLGRYRLVRFEDLAREPLAIBIRALYFTGLTLPQL 300
Db 240 RVREVCRSVRIAEATLKPPLGRYRLVRFEDLAREPLAIBIRALYFTGLTLPQL 299
QY 301 EAWIHNTGSGIGKPIEAFHTSSRNARNVSOQWRHALPFTKILRVOECAGALQLLGYR 360
Db 300 EAWIHNTGSGIGKPIEAFHTSSRNARNVSOQWRHALPFTKILRVOECAGALQLLGYR 359
QY 361 PVYSADOQDRLTDLVLPGRPDHFSWAS 388
Db 360 PVYSEDEQRNALDVLPRGLNGFTWAS 387

RESULT 6
ABB81554
ID ABB81554 standard; protein; 395 AA.
XX AC ABB81554;
XX DT 05-SEP-2002 (first entry)
XX DE Human corneal N-acetylglucosamine-6-sulfotransferase SEQ ID NO:2.
XX KW Human; N-acetylglucosamine-6-sulfotransferase; enzyme; GlcNAc6ST;
XX KW corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
XX KW ophthalmological; chromosome 16q22.
XX OS Homo sapiens.
XX PN US2002061562-A1.
XX PD 23-MAY-2002.
XX PF 09-AUG-2001; 2001US-00927602.
XX PR 11-AUG-2000; 2000US-00638211.
XX PR 11-AUG-2000; 2000US-0325773P.
XX PA (FUKU//) FUKUDA M N.
XX PA (AKAM//) AKAMA T O.
XX PI Fukuda MN, Akama TO;
XX DR WPI; 2002-507643/54.
XX DR N-PSDB; ABN89506.
XX PT New nucleic acid encoding corneal N-acetylglucosamine-6-sulfotransferase,
XX PT useful for treatment, monitoring and diagnosis of macular corneal
XX PT dystrophy.
XX PS Claim 13; Fig 1A-D; 69pp; English.
XX CC The present sequence represents human corneal N-acetylglucosamine-6-
XX CC sulfotransferase (I), which is able to catalyze sulfation of keratan
XX CC sulfate (KS). Also described is a method for monitoring the effect of
XX CC treatments for macular corneal dystrophy (MCD), and detecting
XX CC susceptibility to MCD. (I) is located to chromosome 16q22, and has
XX CC ophthalmological activity. (I) can be used to treat or prevent macular
XX CC corneal dystrophy types I or II. (I) makes possible treatment of MCD

CC without requiring keratinoplasty or keratectomy
XX SQ Sequence 395 AA;
Query Match 84.3%; Score 1729.5; DB 5; Length 395;
Best Local Similarity 85.8%; Pred. No. 2.4e-180;
Matches 333; Conservative 17; Mismatches 37; Indels 1; Gaps 1;
QY 1 MWLPRVSSKTVTVLLLAQTTCLLFIISRPSPSPAGGEDRVHVVLSSWRSGSSFVQL 60
Db 1 MWLPRVSSKTVTVLLLAQTTCLLFIISRPSPSPAGGEDRVHVVLSSWRSGSSFVQL 59
QY 61 FSOHPDVYLMPEPAHWVTTLSQGSAAATLHMAVRDLMSIFLCMDVDFDAYMPOSRLSA 120
Db 60 FNQHPDVYLMPEPAHWVTTLSQGSAAATLHMAVRDLVRSVFLCMDVDFDAYLPWRRNLSD 119
QY 121 FFWNATSRALCSPSPACSAFPRGTISKQDVCKTLCTROPFSLAREACRSYSHVVLKEVRF 180
Db 120 LFQWAVSRALCSPSPACSAFPRGAISSEAVCKPLCARQSFTLAREACRSYSHVVLKEVRF 179
QY 181 NLQVLYPLSDPALNLRIHVLRDPRAVLRSPAGGEARVHVVLSSWRSGSSFVQL 240
Db 180 NLQVLYPLSDPALNLRIHVLRDPRAVLRSPAGGEARVHVVLSSWRSGSSFVQL 239
QY 241 RLIREVCRSHVRIAEATLKPPLGRYRLVRFEDLAREPLAIBIRALYFTGLTLPQL 300
Db 240 RVREVCRSVRIAEATLKPPLGRYRLVRFEDLAREPLAIBIRALYFTGLTLPQL 299
QY 301 EAWIHNTGSGIGKPIEAFHTSSRNARNVSOQWRHALPFTKILRVOECAGALQLLGYR 360
Db 300 EAWIHNTGSGIGKPIEAFHTSSRNARNVSOQWRHALPFTKILRVOECAGALQLLGYR 359
QY 361 PVYSADOQDRLTDLVLPGRPDHFSWAS 388
Db 360 PVYSEDEQRNALDVLPRGLNGFTWAS 387

RESULT 7
AAE15438
ID AAE15438 standard; protein; 395 AA.
XX AC AAE15438;
XX DT 12-MAR-2002 (first entry)
XX DE Human drug metabolising enzyme (DME)-5.
XX KW Human; drug metabolising enzyme; gene therapy; autoimmune disorder;
XX KW inflammatory disorder; acquired immune deficiency syndrome; infection;
XX KW AIDS; adult respiratory distress syndrome; cell proliferative disorder;
XX KW allergy; anaemia; conjunctivitis; actinic keratosis; arteriosclerosis;
XX KW cancer; endocrine disorder; hypothalamus disorder; pituitary disorder;
XX KW gastrointestinal disorder; metabolic disorder; developmental disorder;
XX KW liver disorder; iritis; cystic fibrosis; Addison's disease; reinitis;
XX KW goitre; diabetes; dysphagia; gastric carcinoma; anorexia; eye disorder;
XX KW DME-5.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT Peptide 1..35 /label= Signal_peptide
XX FT Peptide 1..32 /label= Signal_peptide
XX FT Protein 33..395 /note= "Human mature DME-5 protein"
XX FT Protein 36..395 /note= "Human mature DME-5 protein"
XX WO200179468-A2.
XX 25-OCT-2001.

PF 12-APR-2001; 2001WO-US011869.
 XX
 PR 13-APR-2000; 2000US-0197590P.
 PR 19-APR-2000; 2000US-0198403P.
 PR 28-APR-2000; 2000US-0200185P.
 PR 05-MAY-2000; 2000US-0202234P.
 PR 11-MAY-2000; 2000US-0203509P.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Policky JL, Hafalia A, Burford N, Ring HZ, Lal P, Tribouley CM;
 PI Yao MG, Yue H, Tang YT, Patterson C, Das D, Sanjanwala MS;
 PI Gandhi AR, Reddy R, Khan FA, Baughn MR, Ramkumar J, Griffin JA;
 PI Au-Young J;
 XX
 DR WPI; 2002-066363/09.
 DR N-PSDB; AAD24670.
 XX
 PT Novel isolated human drug metabolizing enzymes referred as DME 1-10,
 PT useful for diagnosing, treating, or preventing disorders associated with
 PT aberrant expression of DME such as allergy, anemia, asthma, infertility.
 XX
 PS Claim 1a; Page 131-132; 143pp; English.
 XX
 CC The invention relates to human drug metabolising enzymes referred as DME
 CC and nucleic acid molecules encoding such enzymes. Polynucleotides of the
 CC invention are useful for assessing toxicity of test compounds and in gene
 CC therapy. Sequences of the invention are useful in the diagnosis,
 CC prevention and treatment of autoimmune/inflammatory disorders such as
 CC acquired immune deficiency syndrome (AIDS), adult respiratory distress
 CC syndrome, allergies, anaemia, atherosclerosis, asthma, autoimmune
 CC haemolytic anaemia, contact dermatitis, Crohn's disease,
 CC glomerulonephritis, Goodpasture's syndrome, gout, Grave's disease,
 CC Hashimoto's thyroiditis, irritable bowel syndrome, multiple sclerosis,
 CC osteoarthritis, osteoporosis, psoriasis, systemic lupus erythematosus,
 CC rheumatoid arthritis, scleroderma, ulcerative colitis, uveitis, viral,
 CC bacterial, fungal, parasitic, protozoal, helminthic infections; cell
 CC proliferative disorders such as actinic keratosis, arteriosclerosis,
 CC atherosclerosis, Duchenne and Becker dystrophy, hepatitis, Cushing's
 CC syndrome, cancers, myelodysplastic syndrome, epilepsy; endocrine
 CC disorders such as disorders of the hypothalamus and pituitary resulting
 CC from lesions such as primary brain tumours, adenomas, infarction
 CC associated with pregnancy, aneurysms, vascular malformations; eye
 CC disorders such as conjunctivitis, iritis, retinitis, glaucoma, pigmentosa
 CC ; metabolic disorders such as Addison's disease, cystic fibrosis,
 CC diabetes, goitre, glycogen storage diseases, hypercholesterolemia,
 CC hyperthyroidism, hypoglycaemia, lipid myopathies, Menkes syndrome,
 CC mannosidosis, obesity; gastrointestinal disorders such as dysphagia,
 CC gastric carcinoma, anorexia, nausea, gastroenteritis,
 CC hyperbilirubinaemia, emesis, cirrhosis, diarrhoea, jaundice, Reye's
 CC syndrome, peliosis hepatitis, hepatic vein thrombosis and developmental
 CC disorders. The present sequence is human DME-5 protein
 XX
 SQ Sequence 395 AA;
 Query Match 84.3%; Score 1729.5; DB 5; Length 395;
 Best Local Similarity 85.8%; Pred. No. 2.4e-180;
 Matches 333; Conservative 17; Mismatches 37; Indels 1; Gaps 1;
 QY 1 MWLPRFSSKTVTVLLAQTTCLLFIISRPSPAGGEDRVHVLVLSWRSGSFLGQL 60
 DB 1 MWLPRVSTAVTALLAQ-TFLLFLVSRPGSPAGGEARVHVLVLSWRSGSFLGQL 59
 QY 61 FSDHPDVFYLMFPAWHVWTTLSQGSATLHMAVRDLMSRIFLCMDVDFAYMQSRNLSA 120
 DB 60 FNQHPDVFYLMFPAWHVWTTLSQGSATLHMAVRDLVRSVFLCMDVDFAYLPWRNLS 119
 QY 121 FFWNATSRALCSPACSPAPRGITISKQDVCKTLCTQPPSLAREACRSYHVVLEKVRFF 180
 DB 120 LFQWAVSRALCSPACSPAPRGAISSAEVCKPCARQSFLLAREACRSYHVVLEKVRFF 179
 QY 181 NLQVLYPLSDPALNLRIVHLVRDPRAVLRSRBAAGPILARDNGIVLGTNGKWEADPHL 240
 XX

DB 180 NLQVLYPLSDPALNLRIVHLVRDPRAVLRSRBAAGPILARDNGIVLGTNGKWEADPHL 239
 QY 241 RLIREVCRSHVRIAEAAATLKPPPEFLRGYRLVRFEDLAREPLAEIRALYAFTGLTTPQL 300
 DB 240 RVREVCVRSHVRIAEAAATLKPPPEFLRGYRLVRFEDLAREPLAEIRALYAFTGLTTPQL 299
 QY 301 EAWIHNIHTGSGIGKPIEAFHTSSRNARNVSAWRHALPFTTKILRVQEVCAQALQLLYR 360
 DB 300 EAWIHNIHTGSGIGKPIEAFHTSSRNARNVSAWRHALPFTTKILRVQEVCAQALQLLYR 359
 QY 361 PYSADQQRDLTDLVLPGRPDHFSWAS 388
 DB 360 PYSDEQENLADLDLVLPRGLNGFTWAS 387
 RESULT 8
 ADI21086
 ID ADI21086 standard; protein; 395 AA.
 XX
 AC ADI21086;
 XX
 DT 15-APR-2004 (first entry)
 XX
 DE Novel human protein #61.
 XX
 KW forensic; nutritional source; damaged tissue; diseased tissue;
 KW myeloid cell disorder; lymphoid cell disorder;
 KW bone cartilage tissue growth; tendon tissue growth;
 KW ligament tissue growth; nerve tissue growth; regeneration; wound healing;
 KW tissue repair; tissue replacement; burn; incision; ulcer; cancer; human.
 XX
 OS Homo sapiens.
 XX
 PN WO2003025148-A2.
 XX
 PD 27-MAR-2003.
 XX
 PF 19-SEP-2002; 2002WO-US029964.
 PR 19-SEP-2001; 2001US-0323739P.
 PR 13-SEP-2002; 2002US-00323739.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;
 PI Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang D;
 PI Haley-Vicente D;
 XX
 DR WPI; 2003-354603/33.
 DR N-PSDB; ADI21802.
 XX
 PT New polynucleotides and secreted proteins, useful for treating myeloid or
 PT lymphoid cell disorders, in bone cartilage, tendon, ligament and nerve
 PT tissue growth or regeneration, in wound healing, and in tissue repair and
 PT replacement.
 XX
 PS Claim 20; SEQ ID NO 337; 156pp; English.
 XX
 CC The invention relates to an isolated polynucleotide encoding a
 CC polypeptide with biological activity. The polynucleotides and
 CC polypeptides are useful in diagnostics, forensics, gene mapping,
 CC identification of mutations responsible for genetic disorders and other
 CC traits, to assess biodiversity, as nutritional sources or supplements.
 CC The polynucleotides may also be used as molecular weight markers,
 CC chromosome markers or map related gene positions, or as an antigen to
 CC raise anti-DNA antibodies or elicit immune response. The polypeptides are
 CC useful for raising antibodies, as markers for tissues in which the
 CC corresponding polypeptide is expressed, for re-engineering damaged or
 CC diseased tissues, for treating myeloid or lymphoid cell disorders, in
 CC bone cartilage, tendon, ligament and/or nerve tissue growth or
 CC regeneration, in wound healing, in tissue repair and replacement, in
 CC healing of burns, incisions and ulcers, and in treating cancer. The
 CC present sequence represents the amino acid sequence of a novel human

CC protein.
 XX
 SQ Sequence 395 AA;

Query Match 84.3%; Score 1729.5; DB 7; Length 395;
 Best Local Similarity 85.8%; Pred. No. 2.4e-180; Indels 1; Gaps 1;
 Matches 333; Conservative 17; Mismatches 37;

QY 1 MWLPRFSSKTVTVLLLAQTCLLLFIISRPSPAGGEDRVHVLVLSWRSGSFFLQQL 60
 DB 1 MWLPRVSSSTAVTALLLAQ-TFLLLVLSRPGSPAGGEARVHVLVLSWRSGSFFVQQL 59

QY 61 FSQHPDVFLMEPAWHVWTTLSQSSAATLHMAVRDLMSIFLCMDVDFDAYMPOSRLSA 120
 DB 60 FNQHPDVFLMEPAWHVWTTLSQSSAATLHMAVRDLVRSVFLCDMDVDFDAYLPWRRNLS 119

QY 121 FFWNATSRALCSPACSAFPRGTISKQDVCKTCTROPFSIAREACRSYSHVVLKEVRF 180
 DB 120 LFQWAVSRALCSPACSAFPRGAISSBAVCKPLCARQSFTLAREACRSYSHVVLKEVRF 179

QY 181 NLQVLYPLSDPALNLRIVHLVRDPRAVLRSREAAGPILARDNGIVLGTNGKWEADPHL 240
 DB 180 NLQVLYPLSDPALNLRIVHLVRDPRAVLRSREQTAKALARDNGIVLGTNGTWEADPGL 239

QY 241 RLIREVCRSHVRIAEAAATLKPPLRGYRLVRFEDLAREPLAEIRALYAFTGLTLPQL 300
 DB 240 RVREVCRSHVRIAEAAATLKPPLRGYRLVRFEDLAREPLAEIRALYAFTGLTLPQL 299

QY 301 EAMHNTHTSGGIGKPIEAFTSSRNARNVSOQWRHALPFTKILRVOEVCAGALQLLGYR 360
 DB 300 EAMHNTHTSGGIGKPIEAFTSSRNARNVSOQWRHALPFAKIRRVQELCAGALQLLGYR 359

QY 361 PVYSADOQDRLTLDLVLPGRGDHFSWAS 388
 DB 360 PVYSEDEQRNALDLVLPRLGNGFTWAS 387

RESULT 9
 ADL61235
 ID ADL61235 standard; protein; 395 AA.
 XX
 AC ADL61235;
 XX
 DT 03-JUN-2004 (first entry)
 XX
 DE Human tyrosine kinase biomarker carbohydrate sulphotransferase 6 protein.
 XX
 KW predictor set; protein tyrosine kinase; cytostatic; antiangiogenic;
 KW vasotrophic; vulnary; pharmacogenomic; drug sensitivity; breast cancer;
 KW hypervascular disease; angiogenesis; wound healing scar; human;
 KW biomarker; carbohydrate sulphotransferase 6; enzyme.
 XX
 OS Homo sapiens.
 XX
 PN WO2004020583-A2.
 XX
 PD 11-MAR-2004.
 XX
 PF 26-AUG-2003; 2003WO-US026491.
 XX
 PR 27-AUG-2002; 2002US-0406385P.
 XX
 PA (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX
 PI Huang F, Han X, Reeves KA, Amler L, Fairchild CR, Lee FY;
 PI Shaw P;
 XX
 DR WPI; 2004-239171/22.
 DR N-PSDB; ADL61098.
 XX

New predictor sets with a plurality of polynucleotides and/or
 polypeptides whose expression pattern predicts cell response to a
 compound that modulates protein tyrosine kinase activity, useful in

PT treating breast cancer.
 XX
 PS Claim 9; SEQ ID NO 159; 649pp; English.
 XX

The invention relates to a novel predictor set comprising a plurality of
 polynucleotides and/or polypeptides whose expression pattern is
 predictive of the response of cells to treatment with a compound that
 modulates protein tyrosine kinase activity or members of the protein
 tyrosine kinase pathway. The molecules of the invention demonstrate
 cytosolic, antiangiogenic, vasotrophic and vulnary activities and may
 be useful in the field of pharmacogenomics, in particular for determining
 drug sensitivity and in treating breast cancer, hypervascular diseases,
 angiogenesis and scars in wound healing. The current sequence is that of
 a human protein tyrosine kinase biomarker protein of the invention.
 XX
 SQ Sequence 395 AA;

Query Match 84.3%; Score 1729.5; DB 8; Length 395;
 Best Local Similarity 85.8%; Pred. No. 2.4e-180; Indels 1; Gaps 1;
 Matches 333; Conservative 17; Mismatches 37;

QY 1 MWLPRFSSKTVTVLLLAQTCLLLFIISRPSPAGGEDRVHVLVLSWRSGSFFLQQL 60
 DB 1 MWLPRVSSSTAVTALLLAQ-TFLLLVLSRPGSPAGGEARVHVLVLSWRSGSFFVQQL 59

QY 61 FSQHPDVFLMEPAWHVWTTLSQSSAATLHMAVRDLMSIFLCMDVDFDAYMPOSRLSA 120
 DB 60 FNQHPDVFLMEPAWHVWTTLSQSSAATLHMAVRDLVRSVFLCDMDVDFDAYLPWRRNLS 119

QY 121 FFWNATSRALCSPACSAFPRGTISKQDVCKTCTROPFSIAREACRSYSHVVLKEVRF 180
 DB 120 LFQWAVSRALCSPACSAFPRGAISSBAVCKPLCARQSFTLAREACRSYSHVVLKEVRF 179

QY 181 NLQVLYPLSDPALNLRIVHLVRDPRAVLRSREAAGPILARDNGIVLGTNGKWEADPHL 240
 DB 180 NLQVLYPLSDPALNLRIVHLVRDPRAVLRSREQTAKALARDNGIVLGTNGTWEADPGL 239

QY 241 RLIREVCRSHVRIAEAAATLKPPLRGYRLVRFEDLAREPLAEIRALYAFTGLTLPQL 300
 DB 240 RVREVCRSHVRIAEAAATLKPPLRGYRLVRFEDLAREPLAEIRALYAFTGLTLPQL 299

QY 301 EAMHNTHTSGGIGKPIEAFTSSRNARNVSOQWRHALPFTKILRVOEVCAGALQLLGYR 360
 DB 300 EAMHNTHTSGGIGKPIEAFTSSRNARNVSOQWRHALPFAKIRRVQELCAGALQLLGYR 359

QY 361 PVYSADOQDRLTLDLVLPGRGDHFSWAS 388
 DB 360 PVYSEDEQRNALDLVLPRLGNGFTWAS 387

RESULT 10
 AAY72638
 ID AAY72638 standard; protein; 395 AA.
 XX
 AC AAY72638;
 XX
 DT 02-MAY-2001 (first entry)
 XX
 DE Mouse glycosyl sulfotransferase-4 (GST-4).
 XX
 KW Mouse; glycosyl sulfotransferase-4; GST-4; immunosuppressive; therapy;
 KW selective binding inhibitor; gene therapy; inflammation;
 KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
 KW polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
 KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenailtis;
 KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
 KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
 KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
 KW asthma; hypersensitivity; rheumatic fever; tissue rejection;
 XX
 OS chromosome 851.
 XX
 PT Mus musculus.
 XX

PN WO200106015-A1.
 XX 25-JAN-2001.
 XX 19-JUL-2000; 2000WO-US019741.
 PF 20-JUL-1999; 99US-0144694P.
 PR 13-JUN-2000; 2000US-00593828.
 XX (REGC) UNIV CALIFORNIA.
 XX Rosen SD, Lee JK, Hemmerich S;
 XX WPI; 2001-138471/14.
 DR N-PSDB; AAD02696.
 XX New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
 PT diagnostic and therapeutic agent screening applications.
 XX Claim 3; Fig 2; 128pp; English.
 XX The present sequence is mouse glycosyl sulfotransferase-4 (GST-4). GST-4
 CC gene is found on chromosome 8E1. GST is a type 2 membrane protein useful
 CC for inhibiting a binding event between a selectin and a selectin ligand,
 CC which comprises contacting the selectin with a non-sulphated selectin
 CC ligand, GST and a small molecular agent that inhibits the sulphation
 CC activity of GST. GST is also useful in inhibiting a selectin mediated
 CC binding event. GST is useful in gene therapy to treat disorders such as
 CC acute or chronic inflammation, systemic lupus erythematosus (SLE),
 CC rheumatoid arthritis, polyarteritis nodosa, polymyositis,
 CC dermatomyositis, systemic sclerosis, diabetes, glomerulonephritis,
 CC myasthenia gravis, Sjogren's syndrome, Hashimoto's disease, Grave's
 CC disease, adenitis, hypoparathyroidism, pernicious anaemia,
 CC demyelinating diseases, cirrhosis, ulcerative colitis, dermatitis,
 CC myocarditis, regional enteritis, adult respiratory distress syndrome,
 CC infantile eczema, psoriasis lichen planus, allergic rhinitis, bronchial
 CC asthma, hypersensitivity, rheumatic fever and tissue rejection during
 CC transplantation
 XX Sequence 395 AA;
 SQ

Query Match 74.9%; Score 1536.5; DB 4; Length 395;
 Best Local Similarity 76.0%; Pred. No. 3.6e-159;
 Matches 295; Conservative 30; Mismatches 62; Indels 1; Gaps 1;

Qy 1 MWLPRFSKTVTVLLAQTCLLLFIISRPGSPAGGEDRVHVLVLSWRSGSFFVQL 60
 Db 1 MRLPRFSSTVMLSLLMVQTG-ILVFLVSRQVPSPAGLGRVHVLVLSWRSGSFFVQL 59

Qy 61 FSOHPDVYLMPEAWHWYTTLSQSSAATLHMAVRDLMRSTFLCDMDVDFDAYMQSRNL 120
 Db 60 FSOHPDVYLMPEAWHWYTTLSQSSAATLHMAVRDLMRSTFLCDMDVDFDAYMQSRNL 119

Qy 121 FFWNATSRALCSPACSAFPRGTISKDVCKTLCTROFFSLAREACRSYSHVVLKEVRFF 180
 Db 120 LFWQAVSRALCSPVCEAFARGNISSEVCKPLCATRPFGLAQEACSSYSHVVLKEVRFF 179

Qy 181 NLQVLYPLSDPALNLRIVHVRDPRVAVLSREMAAGPILARDNGIVLGTGKWEADPHL 240
 Db 180 NLQVLYPLSDPALNLRIVHVRDPRVAVLSREMAAGPILARDNGIVLGTGKWEADPHL 239

Qy 241 RLIREVCRSHVRAEATLKPPLRCGRYRLVRFEDLAREPLAEIRALYFTGLTLPQL 300
 Db 240 RVNEVCRSHVRAEALHKKPPPLQDRYRLVRFEDLAREPLAEIRALYFTGLTLPQL 299

Qy 301 EAWIHNITHSGIGKPIEAFHTISRNARNYSQAWRHALPFTKILRVOEVACAGALQLGYR 360
 Db 300 QTIWNIHITHSGPGARAEAKFTTSRDALSVSQAWRHLPFAKIRRVQELCGGALQLGYR 359

Qy 361 PVTYADQORDLTDLVLPGRGDFHFWAS 388
 Db 360 SVHSELEQRDLSDLLPRGMDSPFKWAS 387

RESULT 11
 AAU11275
 ID AAU11275 standard; protein; 395 AA.
 XX AC AAU11275;
 XX 12-MAR-2002 (first entry)
 XX Murine intestinal-GlcNAc 6-sulfotransferase (I-GlcNAc6ST) protein.
 XX Mouse; beta1,3GnT; beta1,3-N-acetylglucosaminyltransferase; MECA-79;
 KW L-selectin; L-selectin sulfotransferase-2; Crohn's disease; diabetes;
 KW ulcerative colitis; inflammatory skin disorder; psoriasis; lichen planus;
 KW allergic contact dermatitis; lymphoma; chronic pneumonia; LSST-2;
 KW delayed-type hypersensitivity reaction; hyperplastic thymus; antiulcer;
 KW antiinflammatory; antipsoriatic; antidiabetic; dermatological;
 KW antiallergic; intestinal-GlcNAc 6-sulfotransferase; I-GlcNAc6ST.
 XX OS Mus musculus.
 XX WO200185177-A1.
 PN 15-NOV-2001.
 PD 10-MAY-2001; 2001WO-US015452.
 PF 11-MAY-2000; 2000US-00569320.
 PR (BURN-) BURNHAM INST.
 XX Fukuda M, Yeh J, Hiraoka N;
 XX WPI; 2002-075226/10.
 DR N-PSDB; AAS16948.
 XX New enzyme, useful for modifying acceptor molecule, comprises an isolated
 PT L-selectin sulfotransferase-2 that directs expression of L-selectin
 PT ligand antigen, MECA-79 in Chinese hamster ovary cells, or intestinal
 PT GlcNAc 6-sulfotransferase.
 XX Claim 28; Fig 10; 98pp; English.
 XX The present invention provides a method of modifying an acceptor molecule
 CC by contacting the acceptor with an isolated beta1,3-N-
 CC acetylglucosaminyltransferase (beta1,3GnT) or an active fragment, where
 CC beta1,3GnT directs expression of a MECA-79 antigen. The invention also
 CC provides a method of treating or preventing an L-selectin-mediated
 CC condition by reducing the expression or activity of a beta1,3GnT that
 CC directs expression of a MECA-79 antigen. This can be done by
 CC administering to the subject an oligosaccharide L-selectin antagonist
 CC that inhibits binding of L-selectin to a MECA-79 antigen, for example by
 CC administering antibody material that specifically binds beta1,3GnT,
 CC and/or a beta1,3GnT antisense nucleic acid molecule. L-selectin
 CC sulfotransferase-2 (LSST-2) also directs MECA-79 antigen expression.
 CC Alternatively, the expression or activity of LSST-2 or its active
 CC fragment can be reduced in combination with reducing the expression or
 CC activity of beta1,3GnT. The method is useful for treating L-selectin
 CC mediated conditions such as Crohn's disease and ulcerative colitis,
 CC inflammatory disorders of the skin such as allergic contact dermatitis,
 CC psoriasis and lichen planus, lymphomas, chronic pneumonia, delayed-type
 CC hypersensitivity reactions, diabetes and hyperplastic thymus. This
 CC sequence represents mouse I-GlcNAc6ST
 XX Sequence 395 AA;
 SQ

Query Match 74.9%; Score 1536.5; DB 5; Length 395;
 Best Local Similarity 76.0%; Pred. No. 3.6e-159;
 Matches 295; Conservative 30; Mismatches 62; Indels 1; Gaps 1;

Qy 1 MWLPRFSKTVTVLLAQTCLLLFIISRPGSPAGGEDRVHVLVLSWRSGSFFVQL 60
 Db 1 MRLPRFSSTVMLSLLMVQTG-ILVFLVSRQVPSPAGLGRVHVLVLSWRSGSFFVQL 59

QY 61 FSQHPDVLYLMEPAHWVWTTLSQGSAAATLHMAVRDLMSIFLCMDVDFDAYMPOSRLSA 120
DB 60 FSQHPDVLYLMEPAHWVWTTLSQGSAAATLHMAVRDLMSIFLCMDVDFDAYMPOSRLSA 119
QY 121 FFWATSRALCSPACAPRGTTISQDVCKTCTROPFSLAREACRSYSHVVLKEVRFF 180
DB 120 LFQWAVSRALCSPVCEAFARGNISSEVCKPLCATRPFGLAQEACSSYSHVVLKEVRFF 179
QY 181 NLQVLYPLSDPALNLRIVHLRDPRAVLSRRAAGPILARDNGIVLGTNGKWEADPHL 240
DB 180 NLQVLYPLSDPALNLRIVHLRDPRAVLSRREGTAKALARDNGIVLGTNGTWVEADPRL 239
QY 241 RLIREVCRSHVRIAEAAATLKPPPLRGYRLVRPDLAREPLAEIRALYFTGLTLPQL 300
DB 240 RVNEVCRSHVRIAEAAATLKPPPLRGYRLVRPDLAREPLAEIRALYFTGLTLPQL 299
QY 301 EAWIHNTHGSGIGKPIEAFHTSSRNARNVSQAWRHLPFTKILRVOECAGALQLLGYR 360
DB 300 QTWIHNTHGSGIGKPIEAFHTSSRNARNVSQAWRHLPFTKILRVOECAGALQLLGYR 359
QY 361 PVYSADOORDLTLDLVLPGRGDHFSWAS 388
DB 360 SVHSEQRDLSDLLPLPRGMDSPFKWAS 387

RESULT 12
ABB81557
ID ABB81557 standard; protein; 418 AA.
AC ABB81557;
DT 05-SEP-2002 (first entry)
DE Mouse intestinal N-acetylglucosamine-6-sulfotransferase SEQ ID NO:5.
KW Human; N-acetylglucosamine-6-sulfotransferase; enzyme; GlcNAc6ST;
KW corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
KW ophthalmological.
OS Mus musculus.
FN US2002061562-A1.
PD 23-MAY-2002.
PF 09-AUG-2001; 2001US-00927602.
PR 11-AUG-2000; 2000US-00638211.
PR 11-AUG-2000; 2000US-0325773P.
XX (FUKU/) FUKUDA M N.
XX (AKAM/) AKAMA T O.
XX Fukuda MN, Akama TO;
XX WPI; 2002-507643/54.
XX New nucleic acid encoding corneal N-acetylglucosamine-6-sulfotransferase,
XX useful for treatment, monitoring and diagnosis of macular corneal
XX dystrophy.

Example 5; Page 24-25; 69pp; English.

The present invention describes human corneal N-acetylglucosamine-6-sulfotransferase (I), which is able to catalyze sulfation of keratan sulfate (KS). Also described is a method for monitoring the effect of treatments for macular corneal dystrophy (MCD), and detecting susceptibility to MCD. (I) is located to chromosome 16q22, and has ophthalmological activity. (I) can be used to treat or prevent macular corneal dystrophy types I or II. (I) makes possible treatment of MCD without requiring keratoplasty or keratotomy. The present sequence represents mouse intestinal N-acetylglucosamine-6-sulfotransferase,

CC which is given in comparison with (I) in the exemplification of the
XX present invention
SQ Sequence 418 AA;
Query Match 74.3%; Score 1523; DB 5; Length 418;
Best Local Similarity 75.8%; Pred. No. 1.2e-157;
Matches 294; Conservative 30; Mismatches 62; Indels 2; Gaps 2;
QY 1 MWLPRSSKTKTVLLLAQTCLLFTISRPGPSAGGEDRVHVLVSSWRSGSFLQQL 60
DB 25 MWLPRSSKTKTVLLLAQTCLLFTISRPGPSAGGEDRVHVLVSSWRSGSFLQQL 83
QY 61 FSQHPDVLYLMEPAHWVWTTLSQGSAAATLHMAVRDLMSIFLCMDVDFDAYMPOSRLSA 120
DB 84 FSQHPDVLYLMEPAHWVWTTLSQGSAAATLHMAVRDLMSIFLCMDVDFDAYMPOSRLSA 143
QY 121 FFWATSRALCSPACAPRGTTISQDVCKTCTROPFSLAREACRSYSHVVLKEVRFF 180
DB 144 LFQWAVSRALCSPVCEAFARGNISSEVCKPLCATRPFGLAQEACSSYSHVVLKEVRFF 203
QY 181 NLQVLYPLSDPALNLRIVHLRDPRAVLSRRAAGPILARDNGIVLGTNGKWEADPHL 240
DB 204 NLQVLYPLSDPALNLRIVHLRDPRAVLSRREGTAKALARDNGIVLGTNGTWVEADPRL 263
QY 241 RLIREVCRSHVRIAEAAATLKPPPLRGYRLVRPDLAREPLAEIRALYFTGLTLPQL 300
DB 264 RVNEVCRSHVRIAEAAATLKPPPLRGYRLVRPDLAREPLAEIRALYFTGLTLPQL 322
QY 301 EAWIHNTHGSGIGKPIEAFHTSSRNARNVSQAWRHLPFTKILRVOECAGALQLLGYR 360
DB 323 QTWIHNTHGSGIGKPIEAFHTSSRNARNVSQAWRHLPFTKILRVOECAGALQLLGYR 382
QY 361 PVYSADOORDLTLDLVLPGRGDHFSWAS 388
DB 383 SVHSEQRDLSDLLPLPRGMDSPFKWAS 410

RESULT 13

ADJ70405
ID ADJ70405 standard; protein; 394 AA.
AC ADJ70405;
DT 06-MAY-2004 (first entry)
DE Human heat mitochondrial protein as a therapeutic target SeqID2211.
XX Mitochondrial; human; screening assay; diabetes mellitus;
XX Huntington's disease; osteoarthritis;
XX Leber's hereditary optic neuropathy; LHON;
XX Mitochondrial encephalopathy lactic acidosis and stroke; MELAS;
XX Myoclonic epilepsy ragged red fibre syndrome; MERRF; cancer;
XX Neuroprotective; nontropic; antidiabetic; anticonvulsant; antiarthritic;
XX Osteopathic; ophthalmological; cytostatic.
OS Homo sapiens.
XX WO2003087768-A2.
XX 23-OCT-2003.
XX 04-APR-2003; 2003WO-US010870.
XX 12-APR-2002; 2002US-0372843P.
XX 17-JUN-2002; 2002US-0389987P.
XX 20-SEP-2002; 2002US-0412418P.
XX (MITO-) MITOKOR.
XX (BUCK-) BUCK INST AGE RES.
XX Ghosh SS, Fahy ED, Zhang B, Gibson BW, Taylor SW, Glenn GM;
XX Warnock DB;

XX WPI; 2003-845369/78.
XX
PT Identifying a mitochondrial target for drug screening assays and for
PT treating diseases associated with altered mitochondrial function,
PT comprises detecting a modified polypeptide in a sample and correlating
PT with the disease.
XX
PS Claim 1; SEQ ID NO 2211; 180pp; English.
XX
XX This invention relates to novel mitochondrial targets that can be used
XX for therapeutic intervention in treating a disease associated with
XX altered mitochondrial function. Specifically, it refers to a method for
XX identifying proteins of the human heart mitochondrial proteome that are
XX useful for drug screening assays, as well as therapeutic targets. The
XX present invention describes a method for identifying such proteins that
XX can be used in the treatment of various diseases associated with altered
XX mitochondrial function including diabetes mellitus, Huntington's disease,
XX osteoarthritis, Leber's hereditary optic neuropathy (LHON), mitochondrial
XX encephalopathy lactic acidosis and stroke (MELAS), myoclonic epilepsy
XX ragged red fibre syndrome (MERRF) or cancer. Accordingly, these
XX compositions have neuroprotective, neurotropic, antidiabetic,
XX anticonvulsant, antiarthritic, osteopathic, ophthalmological and
XX cytotatic activities. This polypeptide sequence is a human heart
XX mitochondrial protein of the invention.
XX
SQ Sequence 394 AA;

Query Match 73.8%; Score 1513.5; DB 7; Length 394;
Best Local Similarity 92.6%; Pred. No. 1.2e-156;
Matches 289; Conservative 7; Mismatches 15; Indels 1; Gaps 1;
QY 1 MWLPRESSKTVVLLLAQTCLLLFIISRGSPSPAGGEDRVHVLVLSWSRSGSFLGQL 60
DB 1 MWLPRESSKTVVLLLAQTCLLLFIISRGSPSPAGGEDRVHVLVLSWSRSGSFLGQL 59
QY 61 FSOHPDVFLMEPAWHVWTTLSQSSAATLHMAVRDLMSIFLCMDVDFDAYMPSQRNLSA 120
DB 60 FNQHPDVFLMEPAWHVWTTLSQSSAATLHMAVRDLVRSVFLCMDVDFDAYLWRRNLSL 119
QY 121 FFWNATSRALCSPPPACSAFPRGTISKQDVCKTLCCTROPFSLAREACRSYSHVVLKEVRFF 180
DB 120 LFQWAVSRALCSPPPACSAFPRGTISKQDVCKTLCCTROPFSLAREACRSYSHVVLKEVRFF 179
QY 181 NLQVLYPLLSDDPALNLRIHVLVRDPRAVLRSREAAAGPILARDNGIVLGTNGKVEADPHL 240
DB 180 NLQVLYPLLSDDPALNLRIHVLVRDPRAVLRSREAAAGPILARDNGIVLGTNGKVEADPHL 239
QY 241 RLIREVCRSHVRIAEEATLKPFPFLRGVRLVRFEDLAREPLAEIRALYAFTGLTLPOL 300
DB 240 RLIREVCRSHVRIAEEATLKPFPFLRGVRLVRFEDLAREPLAEIRALYAFTGLTLPOL 299
QY 301 EAWIHNITHGSG 312
DB 300 EAWIHNITHGSG 311

RESULT 14
AAM93309
ID AAM93309 standard; protein; 386 AA.
XX
AC AAM93309;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polypeptide, SEQ ID NO: 2817.
XX
XX Human; full length cDNA; cDNA synthesis; oligo-capping.
XX
OS Homo sapiens.
XX
XX EPI130094-A2.
XX

PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-00114089.
XX
XX 08-JUL-1999; 99JP-00194486.
PR 11-JAN-2000; 2000JP-00118774.
XX
XX 02-MAY-2000; 2000JP-00183765.
PR
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Nishikawa T, Isogai T, Hayaashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
XX WPI; 2001-524255/58.
DR N-PSDB; AAK94229.
XX
XX 830 Primers useful for synthesizing full length cDNA clones and their use
XX in genetic manipulation.
PT
PS Claim 8; SEQ ID NO 2817; 1380pp + Sequence Listing; English.
XX
XX The invention relates to primers for synthesising full length cDNA
XX clones. 830 cDNA molecules encoding a human protein have been isolated
XX and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have
XX been determined. Primers for synthesising the full length cDNA are useful
XX for clarifying the function of the protein encoded by the cDNA. The full
XX length clones were obtained by construction of full length enriched cDNA
XX libraries that were synthesised by the oligo-capping method. The primers
XX enable the production of the full length cDNA easily without any special
XX methods. The present sequence is a polypeptide encoded by a full length
XX human cDNA of the invention. Note: The sequence data for this patent did
XX not form part of the printed specification, but was obtained in CD-ROM
XX format directly from EPO
XX
SQ Sequence 386 AA;

Query Match 49.4%; Score 1013; DB 4; Length 386;
Best Local Similarity 52.5%; Pred. No. 9.9e-102;
Matches 207; Conservative 55; Mismatches 104; Indels 28; Gaps 7;
QY 14 LLLAQTCTLLFIISRP-----GFSPAGGEDRVHVLVLSWSRSGSFLGQ 59
DB 1 MLLPKKMLLLFLVSQMAILALFFHMYSHNISLSMKAPQPMHVLVLSWSRSGSFLGQ 60
QY 60 LFSQHPDVFLMEPAWHVWTTLSQSSAATLHMAVRDLMSIFLCMDVDFDAYM-POSRL 118
DB 61 LFQHPDVFLMEPAWHVWTTLSQSSAATLHMAVRDLIRAVFLCMDVDFDAYMEPPRRQ 120
QY 119 SAFFNATSRALCSPPPACSAFPRGTISKQDVCKTLCCTROPFSLAREACRSYSHVVLKEVR 178
DB 121 SSIQWENSRALCSPACDIIIPQDEIIIPRAHCELLCSQQPFVVEKACRSYSHVVLKEVR 180
QY 179 FFNQLVYPLLSDDPALNLRIHVLVRDPRAVLRSREAAAGPILARDNGIVLGTN-GKWEAD 237
DB 181 FFNQLVYPLLSDDPALNLRIHVLVRDPRAVLRSREAAAGPILARDNGIVLGTN-GKWEAD 240
QY 238 PHLLIREVCRSHVRIAEEATLKPFPFLRGVRLVRFEDLAREPLAEIRALYAFTGLTL 296
DB 241 QPYVNVQVTCQSQLEIYK--TIQSUPKALQERLLVRYEDLAPAPVQAQSRMTFFVGLF 298
QY 297 TPQLEAWIHNITHGSGIGKPIBAFHTSSRNARNVSOARHALPFTTKILRVQEVYACAGALQL 356
DB 299 LPHLQVWVNIITGKMGD--HAFHTNARDALNVSOARWWSLPEYKVSRLQKACGDAMNL 356
QY 357 LGYRVPYSADQQRDLTLDLVLPRGPDHFSWASPD 390
DB 357 LGYRHSRSEQRNLLDLL-----STWTVPE 383

RESULT 15
ADL30784
ID ADL30784 standard; protein; 386 AA.
XX

AC ADL30784;
XX 20-MAY-2004 (first entry)
XX Human protein encoded by a full length cDNA clone seqID 2817.
DE human; medicine; signal transduction; glycoprotein; transcription;
KW oligo-capping method.
XX Homo sapiens.
XX EPI396543-A2.
XX 10-MAR-2004.
XX 07-JUL-2000; 2003EP-00025638.
XX 08-JUL-1999; 99JP-00194486.
PR 11-JAN-2000; 2000JP-00118774.
PR 02-MAY-2000; 2000JP-00183865.
PR 07-JUL-2000; 2000EP-00114089.
XX (REAS-) RES ASSOC BIOTECHNOLOGY.
XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX WPI; 2004-204755/20.
DR N-PSDB; ADL30783.
XX New oligonucleotide primers (830 cDNAs) useful for synthesizing full
PT length human cDNAs.
XX Example 1; SEQ ID NO 2817; 1340pp; English.

XX This invention relates to a novel primers useful for synthesizing full
CC length cDNA molecules that encode human proteins. Specifically, it refers
CC to secretory or membrane proteins that are potential therapeutic agents/
CC target molecules in the field of medicine, and in particular genes
CC encoding proteins that are associated with signal transduction,
CC glycoproteins and transcription. The present invention describes a method
CC for efficiently cloning a full length human cDNA from both the 5' and 3',
CC ends using the oligo-capping method. This polypeptide sequence is a full
XX length human protein of the invention.
SQ Sequence 386 AA;

Query Match 49.4%; Score 1013; DB 8; Length 386;
Best Local Similarity 52.5%; Pred. No. 9.9e-102;
Matches 207; Conservative 55; Mismatches 104; Indels 28; Gaps 7;
Qy 14 LLLAQTCTLLPIISRP-----GPSPPAGGEDRVHVLSSWRSGSSFLGQ 59
Db 1 MLLPKMKLLFLVQSMAILALFPFMYSHNTSSLSMKAPQPMHVLSSWRSGSSFLGQ 60
Qy 60 LPSQHPDVFYLMPEAWHVMWTTLSQSSAATLHMAVRDLMSIFLCMDVDFDAYM-POSRL 118
Db 61 LFGQHPDVFYLMPEAWHVMWTFKSTAWMLHMAVRDLIRAVFLCDMSVDFDAYMEGPRRQ 120
Qy 119 SAFFNWTASRALCSPACAFPRGTISKQVCKTLCTRPFLSAREACRSYSHVVLKEVR 178
Db 121 SSLQWENSRLCSPACADIIIPQDEIIIPRAHCRLLCSQPFVEVVEKACRSYSHVVLKEVR 180
Qy 179 PFNLQVLYPLLSDPALNRIHVLRDPRVLRREAAGPIIARDNGIVLGTN-GKWVAD 237
Db 181 PFNLQVLYPLLSDPALNRIHVLRDPRVLRREAAGPIIARDNGIVLGTN-GKWVAD 240
Qy 238 PHLRLIREVCSHVRVIAEATLKP-PPFLRGYRLVRFEDLAREPLAEIRALYAFGLTL 296
Db 241 QPYVVMQVICSQLEIYK--TIQSLPKALQRYLLVRYEDLARAPVAQTSRMIEFVGLEF 298
Qy 297 TPQLEAWIHNTHTSGIGKPIEAFHTSSRNARNVSOAWRHALPFTKILRVQEVCAQALQL 356

Db 299 LPHLQTVVHNITRGKMGD--HAFHTNARDALNVSOAWRWSLPYKVSRLQKACGDAMNL 356
Qy 357 LGYRPVYSADQQORDLTLDLVLPRGPDHFSWASPD 390
Db 357 LGYRHYRSEQEQRNLLDLL-----STWTVPE 383
Search completed: June 23, 2005, 08:43:23
Job time : 43.1481 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2005, 08:33:13 ; Search time 42.7985 Seconds
(without alignments)
4666.302 Million cell updates/sec

Title: US-10-697-828-8
Perfect score: 2051
Sequence: 1 MWLPFRSSKTVTLVLLAQTT.....LTLDLVLPRGDFHFWASPD 390

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_03:.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|---------------------|
| 1 | 2051 | 100.0 | 390 | 2 | Q7LCN3 homo sapien |
| 2 | 2051 | 100.0 | 411 | 2 | Q9GZS9 homo sapien |
| 3 | 1729.5 | 84.3 | 395 | 2 | Q9GZK3 homo sapien |
| 4 | 1536.5 | 74.9 | 395 | 2 | Q9GZP4 mus musculus |
| 5 | 1119.5 | 54.6 | 392 | 2 | Q9GN39 xenopus lae |
| 6 | 1013 | 49.4 | 386 | 2 | Q9NCG5 homo sapien |
| 7 | 1008 | 49.1 | 386 | 2 | Q9Y5R3 homo sapien |
| 8 | 999 | 48.7 | 370 | 2 | Q9IV46 homo sapien |
| 9 | 965.5 | 47.1 | 388 | 2 | Q9RI11 mus musculus |
| 10 | 958.5 | 46.7 | 388 | 2 | Q9WUE5 mus musculus |
| 11 | 665 | 32.4 | 483 | 2 | Q794G9 mus musculus |
| 12 | 665 | 32.4 | 530 | 2 | O88276 mus musculus |
| 13 | 662 | 32.3 | 530 | 2 | Q80WV3 mus musculus |
| 14 | 659.5 | 32.2 | 483 | 2 | Q9UED5 homo sapien |
| 15 | 659.5 | 32.2 | 530 | 2 | Q9Y4C5 homo sapien |
| 16 | 617.5 | 30.1 | 484 | 2 | Q9EP78 mus musculus |
| 17 | 614.5 | 30.0 | 484 | 2 | Q9N8B0 mus musculus |
| 18 | 610 | 29.7 | 486 | 2 | O75667 homo sapien |
| 19 | 609 | 29.7 | 486 | 2 | Q9NS84 homo sapien |
| 20 | 603 | 29.4 | 485 | 2 | O6XQ68 rattus norv |
| 21 | 579.5 | 28.3 | 479 | 2 | Q7LGC8 homo sapien |
| 22 | 577.5 | 28.2 | 479 | 2 | O75099 homo sapien |
| 23 | 576 | 28.1 | 472 | 2 | O88199 mus musculus |
| 24 | 560 | 27.3 | 474 | 2 | Q9QZL2 rattus norv |
| 25 | 552 | 26.9 | 420 | 2 | O6DBY9 brachydario |
| 26 | 540.5 | 26.4 | 458 | 1 | C6ST CHICK |
| 27 | 520 | 25.4 | 411 | 2 | Q9EQC0 mus musculus |
| 28 | 513.5 | 25.0 | 411 | 2 | O43916 homo sapien |
| 29 | 472.5 | 23.0 | 441 | 2 | O93403 torpedo cal |
| 30 | 403.5 | 19.7 | 257 | 2 | O79415 mus musculus |
| 31 | 353.5 | 17.2 | 225 | 2 | O6FY62 cavia porce |

| | | | | | |
|----|-------|------|------|---|--------|
| 32 | 293.5 | 14.3 | 304 | 2 | Q677T4 |
| 33 | 274 | 13.4 | 486 | 2 | Q9VMC3 |
| 34 | 259.5 | 12.7 | 345 | 2 | Q7QIF9 |
| 35 | 197 | 9.6 | 119 | 2 | Q95JA8 |
| 36 | 174.5 | 8.5 | 363 | 2 | Q9VMC4 |
| 37 | 172 | 8.4 | 120 | 2 | Q9DOK5 |
| 38 | 170.5 | 8.3 | 363 | 2 | Q8MZD1 |
| 39 | 134 | 6.5 | 303 | 2 | Q93JE6 |
| 40 | 119.5 | 5.8 | 1222 | 2 | O81ZU8 |
| 41 | 112 | 5.5 | 199 | 2 | O7V4N3 |
| 42 | 107.5 | 5.2 | 307 | 2 | O92VG4 |
| 43 | 105.5 | 5.1 | 277 | 2 | O68RS5 |
| 44 | 105 | 5.1 | 274 | 2 | Q8PRA0 |
| 45 | 104.5 | 5.1 | 316 | 2 | Q6ZET4 |

ALIGNMENTS

RESULT 1
Q7LCN3 PRELIMINARY; PRT; 390 AA.
AC Q7LCN3
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DB N-acetylglucosamine 6-O-sulfotransferase.
GN Name-I-GlcNAC-6-ST;
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Intestine;
RA MEDLINE=9443499; PubMed=10491328; DOI=10.1006/bbrc.1999.1324;
RX Lee J.-K., Bhakta S., Rosen S.D., Hemmerich S.;
RT "Cloning and characterization of a mammalian N-acetylglucosamine-6-sulfotransferase that is highly restricted to intestinal tissue."
RL Biochem. Biophys. Res. Commun. 263:543-549(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Intestine;
RA Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF176839; AAD56001.1;
DR EMBL; AF176838; AAD56000.1;
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransferase_1; 1.
KW Transferase.
SQ SEQUENCE 390 AA; 43832 MW; 132D1C3B6BE38179 CRC64;

Query Match 100.0%; Score 2051; DB 2; Length 390;
Best Local Similarity 100.0%; Pred. No. 3.3e-175;
Matches 390; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

| | | | |
|----|-----|--|-----|
| Qy | 1 | MWLPFRSSKTVTLVLLAQTTCLLLFIIRPGPSSPAGGEDRVHVLVLSWSGSSFLGQL | 60 |
| Db | 1 | MWLPFRSSKTVTLVLLAQTTCLLLFIIRPGPSSPAGGEDRVHVLVLSWSGSSFLGQL | 60 |
| Qy | 61 | FSQHPDVFYLMPEPAHWYVTTLSQGSAAATLHMAVRLMRSIFCLDMDVFDAYMPQSRNL | 120 |
| Db | 61 | FSQHPDVFYLMPEPAHWYVTTLSQGSAAATLHMAVRLMRSIFCLDMDVFDAYMPQSRNL | 120 |
| Qy | 121 | FFNWATSRALCSPPACSAFPRGTISKQDCKVCTCTROPFSLAREACRSYSHVLKEVRFF | 180 |
| Db | 121 | FFNWATSRALCSPPACSAFPRGTISKQDCKVCTCTROPFSLAREACRSYSHVLKEVRFF | 180 |
| Qy | 181 | NLQVLYELLSDPALNLRIVHLVRDRAVLRSPRAVLRREAGPILARDNGIVLTNGKWEADPHL | 240 |
| Db | 181 | NLQVLYELLSDPALNLRIVHLVRDRAVLRSPRAVLRREAGPILARDNGIVLTNGKWEADPHL | 240 |

```
QY 241 RLIREVCRSHVIAEATLPPPLRGYRLVRFEDLAREPLAEIRALYAFTGLTLPOL 300
Db 241 RLIREVCRSHVIAEATLPPPLRGYRLVRFEDLAREPLAEIRALYAFTGLTLPOL 300
QY 301 EAWIHNTGSGIGKPIEAFHTSSRNARNVSQARHALPFTKILRVOECAGALQLLYR 360
Db 301 EAWIHNTGSGIGKPIEAFHTSSRNARNVSQARHALPFTKILRVOECAGALQLLYR 360
QY 361 PVSADQORDLTLDLVLPRGPDHFSWASPD 390
Db 361 PVSADQORDLTLDLVLPRGPDHFSWASPD 390

RESULT 2
Q9GZS9 PRELIMINARY; PRT; 411 AA.
AC Q9GZS9; Q9UB3;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Intestinal GlcNAc-6-sulfotransferase (Intestinal N-acetylglucosamine-
DE 6-O-sulfotransferase).
GN Name=CHST5;
OS Homo sapiens (Human).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20472330; PubMed=11017086; DOI=10.1038/79987;
RA Akama T.O., Nishida K., Nakayama J., Watanabe H., Ozaki K.,
RA Nakamura T., Doka A., Kawasaki S., Inoue Y., Maeda N., Yamamoto S.,
RA Fujiwara T., Thonar E.J., Shimomura Y., Kinoshita S., Tanigami A.,
RA Fukuda M.N.;
RT "Macular corneal dystrophy type I and type II are caused by distinct
RT mutations in a new sulphotransferase gene.";
RL Nat. Genet. 26:237-241(2000).
DR EMBL; AF246718; AAG28023.1; -.
DR EMBL; AF219991; AAG26326.1; -.
DR Genbank; HGNC:1973; CHST5.
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
KW Transferase.
SQ SEQUENCE 411 AA; 46160 MW; 97642D54BE926E06 CRC64;

Query Match 100.0%; Score 2051; DB 2; Length 411;
Best Local Similarity 100.0%; Pred No. 3.6e-175;
Matches 390; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MWLPFSSKTVTVLLAQTTCLLFIISRPSPSPAGGEDRVHVLSSWRSGSFLGQL 60
Db 22 MWLPFSSKTVTVLLAQTTCLLFIISRPSPSPAGGEDRVHVLSSWRSGSFLGQL 81
QY 61 FSQHPDVFYLMPEAHVWTTLSQSAATLHMVARDLMSIFLCMDVDFDYMVQSRNLSA 120
Db 82 FSQHPDVFYLMPEAHVWTTLSQSAATLHMVARDLMSIFLCMDVDFDYMVQSRNLSA 141
QY 121 FFWNATSRALCSPACSAFPRGTISKQDVCKTCTQPPFLSAREACRSYSHVVLKEVRFF 180
Db 142 FFWNATSRALCSPACSAFPRGTISKQDVCKTCTQPPFLSAREACRSYSHVVLKEVRFF 201
QY 181 NLQVLPLLSDPALNLRIVHLVRDPRAVLRSREAAAGPILARDNGIVLGTNGKWEADPHL 240
Db 202 NLQVLPLLSDPALNLRIVHLVRDPRAVLRSREAAAGPILARDNGIVLGTNGKWEADPHL 261
QY 241 RLIREVCRSHVIAEATLPPPLRGYRLVRFEDLAREPLAEIRALYAFTGLTLPOL 300
Db 262 RLIREVCRSHVIAEATLPPPLRGYRLVRFEDLAREPLAEIRALYAFTGLTLPOL 321
QY 301 EAWIHNTGSGIGKPIEAFHTSSRNARNVSQARHALPFTKILRVOECAGALQLLYR 360
```

```
Db 322 EAWIHNTGSGIGKPIEAFHTSSRNARNVSQARHALPFTKILRVOECAGALQLLYR 381
QY 361 PVSADQORDLTLDLVLPRGPDHFSWASPD 390
Db 382 PVSADQORDLTLDLVLPRGPDHFSWASPD 411

RESULT 3
Q9GZX3 PRELIMINARY; PRT; 395 AA.
AC Q9GZX3;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE N-acetylglucosamine 6-O-sulfotransferase GSI-4beta (Corneal N-
DE acetylglucosamine-6-O-sulfotransferase) (Carbohydrate (N-
DE acetylglucosamine 6-O) sulfotransferase 6).
GN Name=GSI4beta; Synonyms=CHST6;
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21096027; PubMed=11181564; DOI=10.1093/glycob/11.1.75;
RA Hemmerich S., Lee J.-K., Bhakta S., Bistrup A., Ruddle N.R.,
RA Rosen S.D.;
RT "Chromosomal localization and genomic organization of the galactose/N-
RT acetylglucosamine/N-acetylglucosamine 6-O-sulfotransferase gene
RL Glycobiology 11:75-87(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20472330; PubMed=11017086; DOI=10.1038/79987;
RA Akama T.O., Nishida K., Nakayama J., Watanabe H., Ozaki K.,
RA Nakamura T., Doka A., Kawasaki S., Inoue Y., Maeda N., Yamamoto S.,
RA Fujiwara T., Thonar E.J., Shimomura Y., Kinoshita S., Tanigami A.,
RA Fukuda M.N.;
RT "Macular corneal dystrophy type I and type II are caused by distinct
RT mutations in a new sulphotransferase gene.";
RL Nat. Genet. 26:237-241(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Guichwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywicki M.I., Skalek U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RA Director MGC Project;
RA Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF280086; AAG48244.1; -.
DR EMBL; AF219991; AAG26327.1; -.

```

EMBL; BC074883; AAH74883.1; -
 DR EMBL; BC074834; AAH74834.1; -
 DR EMBL; AF219990; AAG26325.1; -
 DR Genew; HGNC.6938; CHST6.
 DR GO; GO:0005794; C:Golgi apparatus; TAS.
 DR GO; GO:0001517; F:N-acetylglucosamine 6-O-sulfotransferase ac. .; TAS.
 DR GO; GO:0006044; P:N-acetylglucosamine metabolism; IC.
 DR InterPro; IPR000863; Sulfotransferase.
 DR Pfam; PF00685; Sulfotransferase.
 KW Transferase.
 SQ SEQUENCE 395 AA; 44098 MW; 433CA60248A48F67 CRC64;
 Query Match 84.3%; Score 1729.5; DB 2; Length 395;
 Best Local Similarity 85.8%; Pred. No. 2.2e-146;
 Matches 333; Conservative 17; Mismatches 37; Indels 1; Gaps 1;
 QY 1 MWLPRFSKTVTVLLLAQTCLLFIISRPSPSPAGGEDRVHVLVLSWSSGSSFLGQL 60
 DB 1 MWLPRFSSTAVTALLAQ-TPLLFLVSRQVSPSPAGGERVHVLVLSWSSGSSFLGQL 59
 QY 61 FSOHPDVFLMEPAWHVWTTLSQSAATLHMVARDLMRSIFLCMDVDFDAYMPOSRLSA 120
 DB 60 FNOHPDVFLMEPAWHVWTTLSQSAATLHMVARDLVRSVFLCDMDVDFDAYLPWRRNLS 119
 QY 121 FFWNATSRALCSPSPACSAFPRGTISKQDVCKTCTCTROPFSLAREACRSYSHVVLKEVRFF 180
 DB 120 LFQWAVSRALCSPSPACSAFPRGAISSSEAVCKPLCARQSFTLAREACRSYSHVVLKEVRFF 179
 QY 181 NLQVLYPLLSDDPALNLRIHVLRDPRVAVLSRREAGPIIARDNGIVLGTNGKWNVEADPHL 240
 DB 180 NLQVLYPLLSDDPALNLRIHVLRDPRVAVLSRREQTAKALARDNGIVLGTNGTWVEADPGL 239
 QY 241 RLIREVCRSHVRIAEATLKPPLRGYRLVRPDLAREPLAERIALYAFGLTGLTPQL 300
 DB 240 RVNVECRSHVRIAEATLKPPLRGYRLVRPDLAREPLAERIALYAFGLTGLTPQL 299
 QY 301 EAWIHNTHSGIGKPIEAFHTSSRNARNVSOAQRHALPFTKILRVOEVCAGALQLLYR 360
 DB 300 EAWIHNTHSGIGKPIEAFHTSSRNARNVSOAQRHALPFAKIRRVQELCAGALQLLYR 359
 QY 361 PVYSADOQRDLTLDLVLPRGDPHFSWAS 388
 DB 360 PVYSEQRNLALDLVLPRLGNGFTWAS 387

RESULT 4
 Q9QUP4 ID Q9QUP4 PRELIMINARY; PRT; 395 AA.
 AC Q9QUP4;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE N-acetylglucosamine 6-O-sulfotransferase.
 GN Name=Chst5; Synonyms=I-GlcNAC-6-ST;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN=C57Bl/6; TISSUE=Intestine;
 RX MEDLINE=99423499; PubMed=10491328; DOI=10.1006/bbrc.1999.1324;
 RX Lee J.K., Bhakta S., Rosen S.D., Hemmerich S.;
 RT "Cloning and characterization of a mammalian N-acetylglucosamine-6-
 sulfotransferase that is highly restricted to intestinal tissue.";
 RL Biochem. Biophys. Res. Commun. 263:543-549 (1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57Bl/6; TISSUE=Intestine;
 RX Lee J.-K., Bhakta S., Rosen S.D., Hemmerich S.;
 RA Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF176841; AAD56003.1; -
 DR EMBL; AF176840; AAD56002.1; -

MGD; MGI:1931825; Chst5.
 DR GO; GO:0005615; C:extracellular space; TAS.
 DR InterPro; IPR000863; Sulfotransferase.
 DR Pfam; PF00685; Sulfotransferase.
 KW Transferase.
 SQ SEQUENCE 395 AA; 44537 MW; 3PDF71E43ED3D383BE CRC64;
 Query Match 74.9%; Score 1536.5; DB 2; Length 395;
 Best Local Similarity 76.0%; Pred. No. 4.5e-129;
 Matches 295; Conservative 30; Mismatches 62; Indels 1; Gaps 1;
 QY 1 MWLPRFSKTVTVLLLAQTCLLFIISRPSPSPAGGEDRVHVLVLSWSSGSSFLGQL 60
 DB 1 MRLPRFSSTVMSLLMVQTG-ILVFLVSRQVSPSPAGGERVHVLVLSWSSGSSFLGQL 59
 QY 61 FSOHPDVFLMEPAWHVWTTLSQSAATLHMVARDLMRSIFLCMDVDFDAYMPOSRLSA 120
 DB 60 FSOHPDVFLMEPAWHVWTTLSQSAATLHMVARDLVRSVFLCDMDVDFDAYLPWRRNLS 119
 QY 121 FFWNATSRALCSPSPACSAFPRGTISKQDVCKTCTCTROPFSLAREACRSYSHVVLKEVRFF 180
 DB 120 LFQWAVSRALCSPSPACSAFPRGAISSSEAVCKPLCATRPFGLAQACRSYSHVVLKEVRFF 179
 QY 181 NLQVLYPLLSDDPALNLRIHVLRDPRVAVLSRREAGPIIARDNGIVLGTNGKWNVEADPHL 240
 DB 180 NLQVLYPLLSDDPALNLRIHVLRDPRVAVLSRREQTAKALARDNGIVLGTNGTWVEADPRL 239
 QY 241 RLIREVCRSHVRIAEATLKPPLRGYRLVRPDLAREPLAERIALYAFGLTGLTPQL 300
 DB 240 RVNVECRSHVRIAEATLKPPLRGYRLVRPDLAREPLAERIALYAFGLTGLTPQL 299
 QY 301 EAWIHNTHSGIGKPIEAFHTSSRNARNVSOAQRHALPFTKILRVOEVCAGALQLLYR 360
 DB 300 QTIWNIHNSGSGPGARAEAFKTSRDALSVSQAWRHLPFAKIRRVQELCAGALQLLYR 359
 QY 361 PVYSADOQRDLTLDLVLPRGDPHFSWAS 388
 DB 360 SVHSEQRDLSDLLLPRLGMDSPFKWAS 387

RESULT 5
 Q6GN39 ID Q6GN39 PRELIMINARY; PRT; 392 AA.
 AC Q6GN39;
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE MGC83048 protein.
 GN Name=MGC83048;
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodinae; Xenopus.
 OC NCBI_TaxID=8355;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP TISSUE=Embryo;
 RX MEDLINE=23388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RX Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RX Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RX Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RX Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Helel P.,
 RX Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RX Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RX Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
 RX Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RX Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RX Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RX Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RX Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RX Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RX Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RX Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,

```
RA Krzywinski M.I., Skalska U., Smalhus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RA Proc. Natl. Acad. Sci. U.S.A. 99:16999-16903(2002).
RN [2]
RN SEQUENCE FROM N.A.
RP TISSUE=Embryo;
RC MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;
RX Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.;
RA "Genetic and genomic tools for Xenopus research: The NIH Xenopus
RT initiative.";
RL Dev. Dyn. 225:384-391(2002).
RN [3]
RN SEQUENCE FROM N.A.
RP TISSUE=Embryo;
RC Klein S., Gerhard D.S.;
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC073681; AAH73681.1; -.
GO; GO:0008146; F:sulfotransferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransferase_1; 1.
SQ SEQUENCE 392 AA; 45969 MW; 7712810F8B4704D7 CRC64;

Query Match 54.6%; Score 1119.5; DB 2; Length 392;
Best Local Similarity 55.6%; Pred. No. 1.1e-91;
Matches 217; Conservative 60; Mismatches 102; Indels 11; Gaps 5;

QY 3 LRFSSKTVTV--LLAQTTCCLLFIISR-----PGSSPAGGEDRVHVLVLSWSSGSS 56
DB 1 MARFRPNVTAVGFLAQTLFLIL-IYSRHTVLPDTETKT--EKVHLLLSWSSGSS 56
QY 57 LQQLFSQHPDVFYLMPEAHVWTTLSQGSAAATHMAVRDLMSIFICDMDFVDPAYMPQSR 116
DB 57 IGOIFSQHPDVFYLMPEAHVWTTLSQGSAAATHMAVRDLMSIFICDMDFVDPAYMPQSR 116
QY 117 NLSAFFNWTALCSPPACSAFPRGTISKQDVCKTLCTROPFSLAREACRSYSHVLUKE 176
DB 117 NVSELFQWAVSRALCSPACSHDFDREAITNETCKIKGNPFKSKEESCNTYSHVLUKE 176
QY 177 VRFNQLVYPLSDPALNLRIVHLVRDPAVLRSAAGPILARDNGIVLGTNGKWEA 236
DB 177 VRFNQLVYPLSDPALNLRIVHLVRDPAVLRSAAGPILARDNGIVLGTNGKWEA 236
QY 237 DPHRLIREVCRSHVRIAEAAATLKPFPFLRGYRLVRFEDLAREPLAEIRALYAFGLTL 296
DB 237 DIRYDMREICQSHAQMYETAMDKPSFLKGRYMLVRYEDVDRDPLREINQMTEFANLKL 295
QY 297 TPQLEAWIHNTHGSGIGKPIEAFTSSRNARVNSQAWRHALPFTKILRVOEVCAGALQL 356
DB 296 TAKLKWFWNITHGVGPGTKKEEFQITSRNAVNSQAWRKDLSFQKIQIKTICKSEMNL 355
QY 357 LGYRPVYSDQQRDLTLDLVRPGDPHFSW 386
DB 356 LGYQFIDSEKERKXMSDFLPRKRYQFSW 385

RESULT 6
Q8NCG5 PRELIMINARY; PRT; 386 AA.
AC Q8NCG5;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DE Hypothetical protein FLJ90265.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RN SEQUENCE FROM N.A.
RA Isogai T., Ota T., Nishikawa T., Hayashi K., Otsuki T., Sugiyama T.,
```

```
RA Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K.,
RA Yamamoto J., Wakanatsu A., Nakamura Y., Kojima S., Nagahara K.,
RA Masuho Y., Ono T., Okano K., Yoshikawa Y., Aotsuka S., Sasaki N.,
RA Hattori A., Okumura K., Iwayanagi T., Ninomiya K.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK074746; BAC11177.1; -.
DR Genes; HGNC:1972; CHST4.
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransferase_1; 1.
SQ SEQUENCE 386 AA; 45160 MW; 861869348319E42A CRC64;

Query Match 49.4%; Score 1013; DB 2; Length 386;
Best Local Similarity 52.5%; Pred. No. 3.7e-82;
Matches 207; Conservative 55; Mismatches 104; Indels 28; Gaps 7;

QY 14 LLLAQTTCCLLFIISR-----GPSSPAGGEDRVHVLVLSWSSGSSFLGQ 59
DB 1 MLLPKKKKLLFLVLSQMAILALFFHMYSHNSSLMSKMQPERMHVLSWSSGSSFLVGQ 60
QY 60 LFSQHPDVFYLMPEAHVWTTLSQGSAAATHMAVRDLMSIFICDMDFVDPAYM-POSRL 118
DB 61 LFSQHPDVFYLMPEAHVWTTLSQGSAAATHMAVRDLMSIFICDMDFVDPAYMPEGPRRQ 120
QY 119 SAFFNWTALCSPPACSAFPRGTISKQDVCKTLCTROPFSLAREACRSYSHVLUKEVR 178
DB 121 SLFQWENSRALCSPACDIIPQDEIIPRAHCRLLCSQQPFVEVVEKACRSYSHVLUKEVR 180
QY 179 FNFNQLVYPLSDPALNLRIVHLVRDPAVLRSAAGPILARDNGIVLGTN-GKWEAD 237
DB 181 FNFNQLVYPLSDPALNLRIVHLVRDPAVLRSAAGPILARDNGIVLGTN-GKWEAD 240
QY 238 PHRLIREVCRSHVRIAEAAATLKPFPFLRGYRLVRFEDLAREPLAEIRALYAFGLTL 296
DB 241 QPYVNVQVICSQLEIYK--TIQSLPKALQERYLLVRYEDLARAFAVQTSRMTEFVGLGF 298
QY 297 TPQLEAWIHNTHGSGIGKPIEAFTSSRNARVNSQAWRHALPFTKILRVOEVCAGALQL 356
DB 299 LPHLQVWVNIITRGKMGD--HAFHTNARDALVNSQAWRSUPEYKVSRLQKACGDMNL 356
QY 357 LGYRPVYSDQQRDLTLDLVRPGDPHFSWASPD 390
DB 357 LGYRHVRSEQRNLLDLL-----STWTYPE 383

RESULT 7
Q9Y5R3 PRELIMINARY; PRT; 386 AA.
AC Q9Y5R3;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE N-acetylglucosamine 6-O-sulfotransferase (L-selectin ligand
DE sulfotransferase GST-3).
GN Name-GST3;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RN SEQUENCE FROM N.A.
RP TISSUE=Tonsil;
RC MEDLINE=99264336; PubMed=10330415; DOI=10.1083/jcb.145.4.899;
RX Bistrup A., Bhakta S., Lee J.K., Belov Y.Y., Gunn M.D., Zuo F.R.,
RA Huang C.C., Kannagi R., Rosen S.D., Hemmerich S.;
RT "Sulfotransferases of two specificities function in the reconstitution
RT of high endothelial cell ligands for L-selectin.";
RL J. Cell Biol. 145:899-910(1999).
RN [2]
RN SEQUENCE FROM N.A.
RP TISSUE=Tonsil;
```

```

RX MEDLINE=21096027; PubMed=11181564; DOI=10.1093/glycob/11.1.1.75;
RA Hemmerich S., Lee J.-K., Bhakta S., Bistrup A., Ruddle N.R.,
RA Rosen S.D.;
RT "Chromosomal localization and genomic organization of the galactose/N-
RT acetylglucosamine/N-acetylglucosamine 6-O-sulfotransferase gene
RL family";
RN Glycobiology 11:75-87(2001).
[3]
RP SEQUENCE FROM N.A.
RX MEDLINE=21332592; PubMed=11439191; DOI=10.1016/S0092-8674(01)00394-4;
RA Yeh J.-C., Hiraoka N., Petryniak B., Nakayama J., Ellices L.G.,
RA Rabuka D., Hindsgaul O., March J.D., Lowe J.B., Fakuda M.;
RT "Novel sulfated lymphocyte homing receptors and their control by a
RT Core1 extension beta 1,3-N-acetylglucosaminyltransferase.";
RL Cell 105:957-969(2001).
DR EMBL; AF131235; AA033015.1; -
DR EMBL; AF280088; AAG48246.1; -
DR EMBL; AF49783; AAK48417.1; -
DR GO; GO:0008146; F:sulfotransferase activity; TAS.
DR GO; GO:0007155; P:cell adhesion; TAS.
DR GO; GO:0006928; P:cell motility; TAS.
DR GO; GO:0007267; P:cell-cell signaling; TAS.
DR GO; GO:0006955; P:immune response; TAS.
DR GO; GO:0006477; P:protein amino acid sulfation; TAS.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
KW Selectin; Transferase.
SQ SEQUENCE 386 AA; 45133 MW; 0C3BB4022417143A CRC64;

Query Match 49.1%; Score 1008; DB 2; Length 386;
Best Local Similarity 52.3%; Pred. No. 1e-81;
Matches 206; Conservative 56; Mismatches 104; Indels 28; Gaps 7;

Qy 14 LLLAQTCLLLFIISRP-----GPSPPAGGEDRVVILVLSWRSGSSFLGQ 59
Db 1 MLLPKMKLLFLVLSQMAILALFFHMYSHNIISSLSMKAQPERMHVLSWRSGSSVFVG 60

Qy 60 LFSQHPDVFYLMPEAWHVTTLTSGSAAATLHMAVYDLMRSIFLCMDVDFDAYM-PQSRNL 118
Db 61 LFGQHPDVFYLMPEAWHVTTFKSTAWMLHMAVYDLMRSIFLCMDVDFDAYMPEGPRQ 120

Qy 119 SAFENWATSRALCSPPACSPFRGTISKQDVCKTLCTRPFLSAREACRSYSHVVLKEVR 178
Db 121 SSLFQWNSRALCSAPACDIIPQDEIIIPRAHCRLLCSQPPFVVEKACRSYSHVVLKEVR 180

Qy 179 PFNLQVLPSSDPALNLRIVHLVRDPAVRVRSRAAGPILARNGIVLGTN-GKWVEAD 237
Db 181 PFNLQSLYPLKDPSLNLHIVHLVRDPAVRVRSRERTKGDLMDSRVIMGQEQKJKEK 240

Qy 238 PHLRLIREVCRSHVRIAEATLKP-PPLRCRYLVRFDLAREPLAEIRALYFTGLTL 296
Db 241 QPYVMQVICSQSEIYK--TIQSLPKALQERYLLVRYEDLARAPVQTSRMFEVGLGF 298

Qy 297 TPQLEAWIHNTHGSGIGKPIEAFHTSSRNARNVSOAWRHAFPTKILRVQEVCAQALQ 356
Db 299 LPHLQVHNITRCKMGD--HAFHTNARDALNVSOAWRWLSLPYKVSRLQKACGDANWL 356

Qy 357 LGYRPVVSADQORDLTLDLVLPRGDPHFWSASPD 390
Db 357 LGYRHRVSRSEQQRNLLDLL-----STWTVPE 383

RESULT 8
Q8IV46 PRELIMINARY; PRT; 370 AA.
AC Q8IV46;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE CHST4 protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Saplinton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16999-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC035282; AAH35282.1; -
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
SQ SEQUENCE 370 AA; 43249 MW; EA8D76EA4E73C625 CRC64;

Query Match 48.7%; Score 999; DB 2; Length 370;
Best Local Similarity 55.9%; Pred. No. 6.4e-81;
Matches 198; Conservative 51; Mismatches 91; Indels 14; Gaps 6;

Qy 40 DRVHLVLSWRSGSSFLGQFSGHPDVFYLMPEAWHVTTLTSGSAAATLHMAVYDLMRS 99
Db 25 ERMHVLVLSWRSGSSFLGQFSGHPDVFYLMPEAWHVTTFKSTAWMLHMAVYDLMRS 84

Qy 100 IFCLDMVDFDAYM-PQSRNL-SAFENWATSRALCSPPACSPFRGTISKQDVCKTLCTRP 158
Db 85 VFLCDMSVDFDAYMEPPRQSSIFQWNSRALCSAPACDIIPQDEIIIPRAHCRLLCSQ 144

Qy 159 FSLAREACRSYSHVVLKEVRFFNLQVLPSSDPALNLRIVHLVRDPAVRVRSRAAGPI 218
Db 145 FEVVEKACRSYSHVVLKEVRFFNLQVLPSSDPALNLRIVHLVRDPAVRVRSRERTKGD 204

Qy 219 LARDNGIVLGTN-GKWVEADPHRLIREVCRSHVRIAEATLKP-PPLRCRYLVRFD 276
Db 205 LMDSRIVMGQEQKJKEKDPYVMQVICSQSEIYK--TIQSLPKALQERYLLVRYED 262

Qy 277 LAREPLAEIRALYFTGLTLTPQLEAWIHNTHGSGIGKPIEAFHTSSRNARNVSOAWRH 336
Db 263 LARAPVQTSRMFEVGLGFLEPHLQVHNITRCKMGD--HAFHTNARDALNVSOAWRW 320

Qy 337 ALPFTKILRVQEVCAQALQQLGYPVVSADQORDLTLDLVLPRGDPHFWSASPD 390
Db 321 SLPYKVSRLQKACGDANWL LGYRHRVSRSEQQRNLLDLL-----STWTVPE 367

RESULT 9
Q9RI11 PRELIMINARY; PRT; 388 AA.
AC Q9RI11;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)
DE L-selectin ligand sulfotransferase (Chst4 protein).

```

GN Name=Chst4;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99361934; PubMed=10435581; DOI=10.1016/S1074-7613(00)80083-7;
RA Hiraoka N., Petryniak B., Nakayama J., Tsuboi S., Suzuki M., Yeh J.C.,
RA Izawa D., Tanaka T., Miyaoka M., Lowe J.B., Fukuda M;
RT "A novel, high endothelial venule-specific sulfotransferase expresses
RT 6-sulfo sialyl Lewis(X), an L-selectin ligand displayed by CD34.";
RL Immunity 11:79-89(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Colon;
RX MEDLINE=22386257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusica K., Farmer A.A., Rubin G.M., Hong L.,
RA Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Vallalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalek U., Smalusz D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Colon;
RA Strausberg R.;
RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF109155; AAD45579.1; -
DR EMBL; BC057886; AAH57886.1; -
DR MGD; MGI:1349479; Chst4.
DR GO; GO:0016021; C:integral to membrane; TAS.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
KW Selectin; Transferrase.
SQ SEQUENCE 388 AA; 44635 MW; 6D5371AFB6884ABE CRC64;

Query Match 47.1%; Score 965.5; DB 2; Length 388;
Best Local Similarity 55.9%; Pred. No. 6.8e-78;
Matches 194; Conservative 41; Mismatches 87; Indels 25; Gaps 6;

QY 42 VHVLLSSWSRSGSFLGQFSQHPDVPYLMPEPAWHVWMTLSQGSAAATLHMAVRLMRSIF 101
DB 42 VHVLLSSWSRSGSFLGQFGQHPDVPYLMPEPAWHVWMTTSTAWKLHMAVRLRSVF 101

QY 102 LCDMDVFDAYM-PQSRNLSPAFFNATSRALCSPPACSAFPRGTISKQDVCKTLCTRPFFS 160
DB 102 LCDMSVFDAYMPPGPRKQSSLFQWEQSRALCSAPVCDFFPAHEISSPKHCKLLCGQPPD 161

QY 161 LAREACBSYSHVVLKEVRFNQLVPLLLSDPALNRLVHLVDRDPRVLRSAAGPILA 220
DB 162 MVEKACSHGFWVLKEVRFNQLVPLLLSDPALNRLVHLVDRDPRVLRSAAGPILA 221

QY 221 RDNGIVLGTNGKWEADPHRLRLIRE-----VCRSHVRIAE-ATLKPPFPFLRGY 269
DB 222 VDSHVLG-----QHLETKEEDQPYAMKICKSQVDIVKAIQTL--PEALQORY 270

QY 270 LRVRFEDLAREPIALRIALYAFVTLTTPLEAWIHNIHSGIGKPIEAFHTSSRNARN 329
DB 270 LRVRFEDLAREPIALRIALYAFVTLTTPLEAWIHNIHSGIGKPIEAFHTSSRNARN 329

DB 271 LFLRYSDLVRAFLAQTRRLYKFVLDLFLPHLQTLVHNVTRGKMGQ--HAFHTNARNALN 328
QY 330 VSOAWEHALPFTKILRLVQVQACAGLQLLCYRVPVYSADQOQDLTDLV 376
DB 329 VSOAWEHALPFTKILRLVQVQACAGLQLLCYRVPVYSADQOQDLTDLV 375
PRT; 388 AA.
ID Q9WUE5 PRELIMINARY;
AC Q9WUE5; TISSUE=Tongue;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 25-OCT-2004 (TREMBLrel. 28, Last annotation update)
DE N-acetylglucosamine 6-O-sulfotransferase (Mus musculus adult male
DE tongue cDNA, RIKEN full-length enriched library, clone:2310003G18
DE product:carbohydrate (Chondroitin 6/keratan) sulfotransferase 4, full
DE insert sequence).
GN Name=Chst4;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6; TISSUE=Tonsil;
RC MEDLINE=99264336; PubMed=10330415; DOI=10.1083/jcb.145.4.899;
RA Bistrup A., Bhakta S., Lee J.K., Belov Y.Y., Gunn M.D., Zuo F.R.,
RA Huang C.C., Kannagi R., Rosen S.D., Hemmerich S.;
RT "Sulfotransferases of two specificities function in the reconstitution
RT of high endothelial cell ligands for L-selectin.";
RL J. Cell Biol. 145:899-910(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Tongue;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Tongue;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Tongue;
RA The FANTOM Consortium;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Tongue;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Tongue;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuunai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Wataniki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,

Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
"RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer";
Genome Res. 10:1757-1771(2000).
[7]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Tongue;
RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA Hanagaki T., Hara A., Hayashizaki N., Hiramoto K., Hiraoka T., Hori F.,
RA Inotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
RA Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
RA Sano H., Sasaki D., Shibata K., Shibata Y., Shingawa A., Shiraki T.,
RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF131236; AAD33016.1; -
DR EMBL; AK009113; BAB26078.1; -
DR MGD; MGI.11349479; Chst4.
DR GO; GO:0016021; C: integral to membrane; TAS.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
KW Transferase.
SQ SEQUENCE 388 AA; 44694 MW; DIE9D7796DP8574D CRC64;

Query Match 46.7%; Score 958.5; DB 2; Length 388;
Best Local Similarity 55.6%; Pred. No. 2.9e-77;
Matches 193; Conservative 42; Mismatches 87; Indels 25; Gaps 6;

QY 42 VHVILVLSWSSGSSFLGQSFQHPDVFYLMPEAWHVWTTLSQGSAAATLHMAVRDLMSIF 101
DB 42 VHVILVLSWSSGSSFLGQSFQHPDVFYLMPEAWHVWTTLSQGSAAATLHMAVRDLMSIF 101

QY 102 LCDMDVFDAYM-POSRLNSAFFNATSRALCSPPACSAFPRGTISKQDVCKLTCTROPFS 160
DB 102 LCDMSVFDAYMFGPRQSSIFQWQSRLCSAPVCDFFPAHEISSPKHCKLGGQFPD 161

QY 161 LARACRSYSHVILKEVRFNQLVYLLSDPALNLRIHVLVDRPRAVLSREAGPILA 220
DB 162 MVEKACRSHGVFLKEVRFNQLVYLLSDPALNLRIHVLVDRPRAVLSREAGPILA 221

QY 221 RDNGLVLTGKWEADPHLRLE-----VCRSHVRIAE-ATLKPPFPFLGRY 269
DB 222 VDSHIVLG-----OHLTIKEEDQPYAMKICKSQVDIVKAIQTL--PEALQRY 270

QY 270 RLVRFDLAREPLAEIRALYAFTGLTLPOLAEIHNITHSGIGKPKIEAFHTSSRNARN 329
DB 271 LFLAYEDLVRAPLAQTTLYKFVGLDFPLHLOTWVYVNRKNGQ--HAFHTNARNALN 328

QY 330 VSOAQRHALPFTKILRVQEVCAQALQLGYRPPVYADQQRDLTLDLV 376
DB 329 VSOAQRWSPVKEVYQLQDACGEMDLLGYLQVRSQEQGNLSLDLL 375

RESULT 11
Q794G9 PRELIMINARY; PRT; 483 AA.
AC Q794G9;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 25-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE N-acetylglucosamine-6-O-sulfotransferase (N-Acetylglucosamine 6-O-sulfotransferase-1).
GN Name=Gn6st-1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.

TISSUE=Whole embryos;
RC MEDLINE=98380482; PubMed=9712885; DOI=10.1074/jbc.273.35.22577;
RA Uchimura K., Muramatsu H., Kadomatsu K., Fan Q., Kurosawa N.,
RA Mitsuoka C., Kannagi R., Habuchi O., Muramatsu T.;
"Molecular cloning and characterization of an N-acetylglucosamine-6-O-sulfotransferase";
J. Biol. Chem. 273:22577-22583(1998).
[2]
SEQUENCE FROM N.A.
RP Uchimura K., Kadomatsu K., El-Pasakhany F.M., Singer M.S., Izawa M.,
RA Kannagi R., Takeda N., Rosen S.D., Muramatsu T.;
"N-Acetylglucosamine 6-O-Sulfotransferase-1 Regulates Expression of L-Selectin Ligands and Lymphocyte Homing";
J. Biol. Chem. 279:35001-35008(2004).
DR EMBL; AB011452; BAA32139.1; -
DR EMBL; AB125058; BAA32139.1; -
DR EMBL; AB011451; BAA32137.1; -
DR GO; GO:0005615; C: extracellular space; TAS.
DR GO; GO:0008146; F: sulfotransferase activity; IDA.
DR GO; GO:0006044; P: N-acetylglucosamine metabolism; TAS.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
KW Transferase.
SQ SEQUENCE 483 AA; 52830 MW; 831FA08F5PFD70E CRC64;

Query Match 32.4%; Score 665; DB 2; Length 483;
Best Local Similarity 40.4%; Pred. No. 7.9e-51;
Matches 150; Conservative 64; Mismatches 121; Indels 36; Gaps 11;

QY 37 GGDRLVHLVLSWSSGSSFLGQSFQHPDVFYLMPEAWHVWTTLSQGSAAATLHMAVRDL 96
DB 113 GGDRLVHLVLSWSSGSSFLGQSFQHPDVFYLMPEAWHVWTTLSQGSAAATLHMAVRDL 172

QY 97 MRSIFLCDMDVFDAYMPO---SRNLS--AFFNATSRALCSPPACSAFPRGTISKQD--V 149
DB 173 LSALYCDLSVFPOLYSPAGSGRNLTITLGFGAATNKKVCCSSFLCPAYRKEVGLVDVRV 232

QY 150 CKTLCTRQPSLAREACRSYSHVILKEVRFNQLVYLLSDPALNLRIHVLVDRPRAVL 209
DB 233 CKK-CPQRLAREECCRYRTVVIKGVFVAVLAPLAKDPALDLKVIHLVDRPRAVA 291

QY 210 RS-----REAGPIIARD-----NGVILGT--NGKWEADPH--LRLE 246
DB 292 SSRIRSHRGLIRESLQVRSRDPRAHRMPPLEAGHKLGAKGCGMGDPADYHALGAMFVI 351

QY 247 CRSHVRIAEATLKPPFPFLGRVRLVRFEDLAREPLAEIRALYAFTGLTLPOLAEIHN 306
DB 352 CNSMAKTLQTA-LQPPDMLOGHVLYVRYEDLVGDPVKTLLRRVDFVGLLVSPENEQFALN 410

QY 307 ITHGSG-IGKPIEAFHTSSRNARNVSOAQRHALPFTKILRVQEVCAQALQLGYRPPVYSA 365
DB 411 MTSGSGSSSKP---FVVSARNATQAANAWRTALTFTQIKQVEEFYQPMALVGYERNVP 467

QY 366 DQORDLTLDLV 376
DB 468 EEVKDLSKTL 478

RESULT 12
O88276 PRELIMINARY; PRT; 530 AA.
ID O88276
AC O88276;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE N-acetylglucosamine-6-O-sulfotransferase long form.
GN Name=Chst2;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.

```
RC TISSUE=Whole embryos;
RX MEDLINE=98380482; PubMed=9712885; DOI=10.1074/jbc.273.35.22577;
RA Uchimura K., Muramatsu H., Kadomatsu K., Pan Q., Kurosawa N.,
RA Mitsuoka C., Kennadi R., Habuchi O., Muramatsu T.;
RT "Molecular cloning and characterization of an N-acetylglucosamine-6-O-
RT sulfotransferase.";
RL J. Biol. Chem. 273:22577-22583 (1998).
DR EMBL; AB011452; BAA32138.1; -.
DR MGD; MGI:1891160; Chst2.
DR GO; GO:0005615; C:extracellular space; TAS.
DR GO; GO:0008146; F:sulfotransferase activity; IDA.
DR GO; GO:0006044; P:N-acetylglucosamine metabolism; TAS.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransferase_1; 1.
KW Transferase
SQ SEQUENCE 530 AA; 57814 MW; A113B1B735C363EC CRC64;

Query Match 32.4%; Score 665; DB 2; Length 530;
Best Local Similarity 40.4%; Pred. No. 8.9e-51;
Matches 150; Conservative 64; Mismatches 121; Indels 36; Gaps 11;

QY 37 GGEDRVHVLVLSWRSGLGQLFSQHPDVFYLMPEAMHVVTTLSQGSAAATLHMAVRDL 96
Db 160 GGDGRQLVYVFTTWRSGLSFFGLFNQNPVEFLYEPVHVWQKLYPGDVAVSLQGAARDM 219
QY 97 MRSIFLCLMDVFDAYMPQ---SRNLS--AFNNWATSRALCSPACSAFPRGTISKOD--V 149
Db 220 LSAlyRCDSLVSQLYSPAGSGGRNLTTLGIFGAATNKVCCSSPLCPAYRKEVVGLVDDRV 279
QY 150 CKTLCTRQPSFLAREACRSYSHVVLKEVRFNQLVPLSDPALNLRIVHLVRDPRAVL 209
Db 280 CKK-CPQRLARFEECRKRYTLVIKGVRFVDAVLAPLLKDPALDKVHLVRDPRAVA 338
QY 210 RS-----REAGPILARD-----NGIVLGT--NGKWVEADPH--LRILREV 246
Db 339 SSRIRSRHGLIRESLQVVRSDRPAHRMPFLEAGHKLGAKGGMGPADYHAGAMEVI 398
QY 247 CRSHVRIAEAAATLKPPPLRGRLVRFEDLAREPLAEIRALYAFTGLTLPQLEAWIHN 306
Db 399 CNSMAKTLQTA-LQPPDWLQGHVLYVRYEDLVGDPVKTLRRVYDFVGLLVSPMEQFALN 457
QY 307 ITHGSG-IGKPIEAFHTSSRNARNVSOAQRHALPFTKILRVQEVCAQALQLLGYRPYISA 365
Db 458 MTSGSGSSSKP---FVVSARNATQAANAWRTALTFOQIKQVEEFQYQPMVGLGVERNVSP 514
QY 366 DQQRDLTLDLV 376
Db 515 EVKDLSTKLL 525

RESULT 13
Q80WV3 ID Q80WV3 PRELIMINARY; PRT; 530 AA.
AC Q80WV3
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Chst2 protein.
GN Name=Chst2;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=98380482; PubMed=9712885; DOI=10.1074/jbc.273.35.22577;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Dege J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Munzy D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettaman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A.C., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalek U., Smailus D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RA "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX Strausberg R.;
RL EMBL; BC051963; AAH51963.1; -.
DR MGD; MGI:1891160; Chst2.
DR GO; GO:0005615; C:extracellular space; TAS.
DR GO; GO:0008146; F:sulfotransferase activity; IDA.
DR GO; GO:0006044; P:N-acetylglucosamine metabolism; TAS.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransferase_1; 1.
KW Transferase
SQ SEQUENCE 530 AA; 57828 MW; 275363BF15440730 CRC64;

Query Match 32.3%; Score 662; DB 2; Length 530;
Best Local Similarity 40.2%; Pred. No. 1.6e-50;
Matches 149; Conservative 65; Mismatches 121; Indels 36; Gaps 11;

QY 37 GGEDRVHVLVLSWRSGLGQLFSQHPDVFYLMPEAMHVVTTLSQGSAAATLHMAVRDL 96
Db 160 GGDGRQLVYVFTTWRSGLSFFGLFNQNPVEFLYEPVHVWQKLYPGDVAVSLQGAARDM 219
QY 97 MRSIFLCLMDVFDAYMPQ---SRNLS--AFNNWATSRALCSPACSAFPRGTISKOD--V 149
Db 220 LSAlyRCDSLVSQLYSPAGSGGRNLTTLGIFGAATNKVCCSSPLCPAYRKEVVGLVDDRV 279
QY 150 CKTLCTRQPSFLAREACRSYSHVVLKEVRFNQLVPLSDPALNLRIVHLVRDPRAVL 209
Db 280 CKK-CPQRLARFEECRKRYTLVIKGVRFVDAVLAPLLKDPALDKVHLVRDPRAVA 338
QY 210 RS-----REAGPILARD-----NGIVLGT--NGKWVEADPH--LRILREV 246
Db 339 SSRIRSRHGLIRESLQVVRSDRPAHRMPFLEAGHKLGAKGGMGPADYHAGAMEVI 398
QY 247 CRSHVRIAEAAATLKPPPLRGRLVRFEDLAREPLAEIRALYAFTGLTLPQLEAWIHN 306
Db 399 CNSMAKTLQTA-LQPPDWLQGHVLYVRYEDLVGDPVKTLRRVYDFVGLLVSPMEQFALN 457
QY 307 ITHGSG-IGKPIEAFHTSSRNARNVSOAQRHALPFTKILRVQEVCAQALQLLGYRPYISA 365
Db 458 MTSGSGSSSKP---FVVSARNATQAANAWRTALTFOQIKQVEEFQYQPMVGLGVERNVSP 514
QY 366 DQQRDLTLDLV 376
Db 515 EVKDLSTKLL 525

RESULT 14
Q9UED5 ID Q9UED5 PRELIMINARY; PRT; 483 AA.
AC Q9UED5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE N-acetylglucosamine-6-O-sulfotransferase (GlcNAc6ST).
GN Name=G6ST;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
```

OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98391845; PubMed=9722682;
RA Uchimura K., Muramatsu H., Kaname T., Ogawa H., Yamakawa T., Pan Q.,
RA Mitsuoka C., Kannagi R., Habuchi O., Yokoyama I., Yamamura K.,
RA Ozaki T., Nakagawara A., Kadamatsu K., Muramatsu T.,
RT "Human N-acetylglucosamine-6-O-sulfotransferase involved in the
RT biosynthesis of 6-sulfo sialyl Lewis X: molecular cloning, chromosomal
RT mapping, and expression in various organs and tumor cells";
RL J. Biochem. 124:670-678(1998).
RN (2)
RN SEQUENCE FROM N.A.
RP TISSUE=Placenta;
RC MEDLINE=20498786; PubMed=11042394; DOI=10.1016/S0304-4165(00)00136-7;
RA Sakaguchi H., Kitagawa H., Sugahara K.,
RT "Functional expression and genomic structure of human N-
RT acetylglucosamine-6-O-sulfotransferase that transfers sulfate to b-N-
RT acetylglucosamine at the nonreducing end of an N-acetylglucosamine
RT sequence";
RL Biochim. Biophys. Acta 1523:269-276(2000).
DR EMBL; AB014679; BAA34265.2; -
DR EMBL; AB021124; BAB16886.1; -
DR EMBL; AB021125; BAB16887.1; -
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransferase_1; 1.
KW Transferase.
SQ SEQUENCE 483 AA; 52787 MW; E27797D44931BA18 CRC64;

Query Match 32.2%; Score 659.5; DB 2; Length 483;
Best Local Similarity 39.1%; Pred. No. 2.4e-50;
Matches 151; Conservative 68; Mismatches 122; Indels 45; Gaps 12;

Qy 31 GPSSPAG-----GDRVHVLVLSWSRSGSFLGOLFQSFDPVFLMEPAWHVMTTL 81
Db 98 GVAAPPNGTRGTGGVGDKQLVVFTTWSRSGSFFGELFNQNPVEVFLYEPVHVWQKL 157

Qy 82 SQGSAATHMAVRDLMSRIFCLDMVDVDPAYMPQ---SRNLS--AFFNWSRALCSPAC 136
Db 158 YPGDAVSLQGAARDMLSAFYCDLSVFLYSPAGSGGRNLTTLGIFGAATNKVVCSSPLC 217

Qy 137 SAFPRGTISKOD--VCKTLCRQPFSLAREACRSYSHVVLKEVFENLVLYPLSDPAL 194
Db 218 PAYRKEVVLVDVDRCKK-CPQRLARFEBCRKYRTLVINGRVFVDAVLAPLRDPAL 276

Qy 195 NLRVHLVRDPAVLRS-----REAAGPILARD-----NGVLGTNGKWV 234
Db 277 DLKVIHLVRDPAVASSRIRSRHGLIRSLQVRSRDPRAHRMPLLEAAGHKLGAKEGV 336

Qy 235 --EADPH-LRLIREVCRSHVRIAEATLKPPFPLRGYRLVRFPDLAREPLAEIRALYAF 291
Db 337 GGPADYHALGAMVEICNSMAKTLQTA-LQPPDWLQGHVLYRVYEDLVGDPVKTLRRVYDF 395

Qy 292 TGLTLTPQLEAWIHNITHGSG-IGKPIEAFTSSRNARNVSOAQRHALPFTKILRVQVQC 350
Db 396 VGLLVSPMEQFALNMTSGSGSSSKP-----FVVSARNATQAAANWRTALTFOQIKQVEEFC 452

Qy 351 AGALQLLGYRPVYSADQORDLTLDIV 376
Db 453 YQPMNAVGLYERVNSPEEVKDLSTLL 478

RESULT 15
Q9Y4C5 PRELIMINARY; PRT; 530 AA.
AC Q9Y4C5; Q9GZNS; Q9Y6F2;
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)
DE Long form of N-acetylglucosamine-6-O-sulfotransferase (GlcNAc6ST)

DE (Carbohydrate sulfotransferase 2).
GN Name=GN6ST; Synonyms=CHST2;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98391845; PubMed=9722682;
RA Uchimura K., Muramatsu H., Kaname T., Ogawa H., Yamakawa T., Pan Q.,
RA Mitsuoka C., Kannagi R., Habuchi O., Yokoyama I., Yamamura K.,
RA Ozaki T., Nakagawara A., Kadamatsu K., Muramatsu T.,
RT "Human N-acetylglucosamine-6-O-sulfotransferase involved in the
RT biosynthesis of 6-sulfo sialyl Lewis X: molecular cloning, chromosomal
RT mapping, and expression in various organs and tumor cells";
RL J. Biochem. 124:670-678(1998).
RN (2)
RN SEQUENCE FROM N.A.
RP TISSUE=Umbilical vein endothelium;
RC MEDLINE=99168906; PubMed=10049591; DOI=10.1006/geno.1998.5653;
RA Li X., Tedder T.F.,
RT "CHST1 and CHST2 sulfotransferases expressed by human vascular
RT endothelial cells: cDNA cloning, expression, and chromosomal
RT localization";
RL Genomics 55:345-347(1999).
DR EMBL; AB014680; BAA34266.2; -
DR EMBL; AF083066; AAD20981.1; -
DR Genew; HGNC:1970; CHST2.
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransferase_1; 1.
KW Transferase.
SQ SEQUENCE 530 AA; 57857 MW; A82CA227B9D5651B CRC64;

Query Match 32.2%; Score 659.5; DB 2; Length 530;
Best Local Similarity 39.1%; Pred. No. 2.8e-50;
Matches 151; Conservative 68; Mismatches 122; Indels 45; Gaps 12;

Qy 31 GPSSPAG-----GDRVHVLVLSWSRSGSFLGOLFQSFDPVFLMEPAWHVMTTL 81
Db 145 GVAAPPNGTRGTGGVGDKQLVVFTTWSRSGSFFGELFNQNPVEVFLYEPVHVWQKL 204

Qy 82 SQGSAATHMAVRDLMSRIFCLDMVDVDPAYMPQ---SRNLS--AFFNWSRALCSPAC 136
Db 205 YPGDAVSLQGAARDMLSAFYCDLSVFLYSPAGSGGRNLTTLGIFGAATNKVVCSSPLC 264

Qy 137 SAFPRGTISKOD--VCKTLCRQPFSLAREACRSYSHVVLKEVFENLVLYPLSDPAL 194
Db 265 PAYRKEVVLVDVDRCKK-CPQRLARFEBCRKYRTLVINGRVFVDAVLAPLRDPAL 323

Qy 195 NLRVHLVRDPAVLRS-----REAAGPILARD-----NGVLGTNGKWV 234
Db 324 DLKVIHLVRDPAVASSRIRSRHGLIRSLQVRSRDPRAHRMPLLEAAGHKLGAKEGV 383

Qy 235 --EADPH-LRLIREVCRSHVRIAEATLKPPFPLRGYRLVRFPDLAREPLAEIRALYAF 291
Db 384 GGPADYHALGAMVEICNSMAKTLQTA-LQPPDWLQGHVLYRVYEDLVGDPVKTLRRVYDF 442

Qy 292 TGLTLTPQLEAWIHNITHGSG-IGKPIEAFTSSRNARNVSOAQRHALPFTKILRVQVQC 350
Db 443 VGLLVSPMEQFALNMTSGSGSSSKP-----FVVSARNATQAAANWRTALTFOQIKQVEEFC 499

Qy 351 AGALQLLGYRPVYSADQORDLTLDIV 376
Db 500 YQPMNAVGLYERVNSPEEVKDLSTLL 525

Search completed: June 23, 2005, 08:48:58
Job time : 43.7985 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2005, 08:33:38 ; Search time 8.8459 Seconds
(without alignments)
4242.027 Million cell updates/sec

Title: US-10-697-828-8

Perfect score: 2051

Sequence: 1 MWLFFRSKTVTVLLLAQTT.....LTLDLVLPRGDFHSWASPD 390

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 79:*

1: Pirl1:*

2: Pirl2:*

3: Pirl3:*

4: Pirl4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----------|---------------------|
| 1 | 659.5 | 32.2 | 484 | 2 JE0261 | N-acetylglucosamin |
| 2 | 613.5 | 29.9 | 484 | 2 JC7350 | N-acetylglucosamin |
| 3 | 610 | 29.7 | 486 | 2 JC7351 | N-acetylglucosamin |
| 4 | 540.5 | 26.4 | 458 | 2 A57397 | chondroitin 6-sulf |
| 5 | 107.5 | 5.2 | 307 | 2 E95934 | probable enzyme, C |
| 6 | 105 | 5.1 | 359 | 2 T16350 | hypothetical prote |
| 7 | 102.5 | 5.0 | 388 | 2 G70729 | hypothetical prote |
| 8 | 96 | 4.7 | 388 | 1 DEHPT | pyruvate dehydroge |
| 9 | 96 | 4.7 | 707 | 2 S29029 | phenylalanine ammo |
| 10 | 95 | 4.6 | 474 | 2 T38905 | probable valine-py |
| 11 | 95 | 4.6 | 885 | 1 S78246 | endopeptidase Clp |
| 12 | 94.5 | 4.6 | 1581 | 1 VGMJBV | peplomer glycoprot |
| 13 | 94 | 4.6 | 582 | 2 A70841 | probable oxalyl-Co |
| 14 | 93.5 | 4.6 | 615 | 2 C75278 | DNA polymerase III |
| 15 | 93 | 4.5 | 523 | 2 B83629 | probable ATP-bindi |
| 16 | 90.5 | 4.4 | 380 | 2 T42755 | tyrosylprotein sul |
| 17 | 88 | 4.3 | 737 | 2 AD2037 | hypothetical prote |
| 18 | 86.5 | 4.2 | 445 | 2 T05639 | hypothetical prote |
| 19 | 86.5 | 4.2 | 577 | 2 B75595 | probable long-chain |
| 20 | 86.5 | 4.2 | 712 | 2 T3231 | hypothetical prote |
| 21 | 86 | 4.2 | 417 | 2 F71238 | hypothetical prote |
| 22 | 85.5 | 4.2 | 221 | 2 T50665 | hypothetical prote |
| 23 | 85.5 | 4.2 | 389 | 1 DSPGPA | pyruvate dehydroge |
| 24 | 85.5 | 4.2 | 425 | 2 E83023 | 3-deoxy-D-manno-Oc |
| 25 | 85.5 | 4.2 | 571 | 2 S69210 | protein kinase cak |
| 26 | 85.5 | 4.2 | 1254 | 2 T41262 | mutS family DNA mi |
| 27 | 85 | 4.1 | 364 | 2 A49781 | cholesterol dehydr |
| 28 | 85 | 4.1 | 696 | 2 A11029 | hypothetical prote |
| 29 | 85 | 4.1 | 776 | 2 T31210 | membrane protein t |

| | | | | | |
|----|------|-----|------|----------|---------------------|
| 30 | 85 | 4.1 | 1182 | 2 S14916 | hypothetical prote |
| 31 | 84.5 | 4.1 | 334 | 2 T44163 | hypothetical prote |
| 32 | 84.5 | 4.1 | 390 | 1 DERTP1 | pyruvate dehydroge |
| 33 | 84.5 | 4.1 | 390 | 1 DERTPA | pyruvate dehydroge |
| 34 | 84.5 | 4.1 | 390 | 2 S23506 | pyruvate dehydroge |
| 35 | 84.5 | 4.1 | 431 | 2 A36036 | cytochrome P450 2F |
| 36 | 84.5 | 4.1 | 1844 | 1 RRWFTM | genome polyprotein |
| 37 | 84 | 4.1 | 552 | 2 T44954 | flagella-related p |
| 38 | 84 | 4.1 | 590 | 2 T39246 | protein phosphotas |
| 39 | 83.5 | 4.1 | 260 | 2 B95965 | probable two-compo |
| 40 | 83.5 | 4.1 | 390 | 1 DEHUPA | pyruvate dehydroge |
| 41 | 83.5 | 4.1 | 420 | 2 A53531 | oncosectal trophobl |
| 42 | 83.5 | 4.1 | 856 | 2 F87316 | conserved hypothet |
| 43 | 83.5 | 4.1 | 910 | 2 E89918 | 2-oxoglutarate deh |
| 44 | 83.5 | 4.1 | 1844 | 2 S01956 | hypothetical prote |
| 45 | 83.5 | 4.1 | 1985 | 2 S19151 | hypothetical prote |

ALIGNMENTS

RESULT 1

JE0261

N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) - human

C:Species: Homo sapiens (man)

C>Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 01-Mar-2002

C:Accession: JE0261

R:Uchimura, K.; Muramatsu, H.; Kaname, T.; Ogawa, H.; Yamakawa, T.; Fan, Q.W.; Mitsuoka, T.

J. Biochem. 124, 670-678, 1998

A:Title: Human N-Acetylglucosamine-6-O-sulfotransferase involved in the biosynthesis of (

A:Reference number: JE0261; MUID:98391845; PMID:9722682

A:Accession: JE0261

A:Molecule type: mRNA

A:Residues: 1-484 <UCH>

A:Cross-references: DDBJ:AB014679

C:Comment: This protein catalyzes the transfer of sulfur of sulfate from 3'-phosphoadenos

C:Superfamily: chondroitin 6-sulfotransferase

C:Keywords: sulfotransferase

Query Match 32.2%; Score 659.5; DB 2; Length 484;
Best Local Similarity 40.0%; Pred. No. 1.5e-51;
Matches 150; Conservative 67; Mismatches 121; Indels 37; Gaps 12;

| | | | |
|----|-----|---|-----|
| Qy | 34 | SPAGGEDRVH-VLVLSSWRSSGFLGOLFQHPDVFVLMPEAHVHTTSLSGSAATHMA | 92 |
| Db | 110 | APEGVGDKRHMMYVFTTWRSGSFFGELFNQNPVEVFLYEPVHWOKLYPGDAVSLQA | 169 |
| Qy | 93 | VRDLMRSIFLCDMDVFDAYMPQ---SRNLS--AFFNWTSRALCSPACSAFPRGTISKQ | 147 |
| Db | 170 | ARDMLSALYKCDLSVFOLYSPAGSGGNLTTLTGIFGATNKVVCSSPLCPAYRKEVGLV | 229 |
| Qy | 148 | D--VKTKLTCTQPFSLAREACRSYSHVVLKEVRFNFNOVLPLSLDPAALNURIHVLRDP | 205 |
| Db | 230 | DDRVCCK-CPQRLARPEECRKYRTLVIKGVRFVDAVLAPLRLDPAALDKVHLVRDP | 288 |
| Qy | 206 | RAVLRS-----REAAGPILARD-----NGVLGTNGKWV--EADPH-LRL | 242 |
| Db | 289 | RAVASSIRSRHGLRESLQVRSRDPRAHMPFLEAGHKLGAKGGVGGPADHALGA | 348 |
| Qy | 243 | IREVCRSHVIAEATLKPPFLRGVRLVRFEOLAREPLAEIRALYAFTLTLPQLEA | 302 |
| Db | 349 | MEVICNSMAKTLQTA-LQPPDWLQGHVLYRYEDLVGEVPKTLRRVDFVGLLSPENEQ | 407 |
| Qy | 303 | WIHNITHGSG-IGKPIBAFTSSRNARNVSOAMRHAPFTTKILRQVEVCAGALQLLVGRP | 361 |
| Db | 408 | FALNMTSGSGSSSKP---FVVSARNATQANAWRTALTFFQIKQVEEFQYQPMVAVLYER | 464 |
| Qy | 362 | VYSADQQRDLTDLV | 376 |
| Db | 465 | VNSPEEVKDLSTLL | 479 |

Db 393 IYSTQKNSQSEKWRFSIPKLAQVQDCEPAMRLFGYKLASSAQBLTNRSL-LEE 451
 Qy 380 GP 381
 Db 452 GP 453

RESULT 5
 E95934
 probable enzyme, C-terminal domain similar to sulfotransferase protein [imported] - Sinc
 C:Species: Sinorhizobium meliloti
 C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 09-Jul-2004
 C:Accession: E95934
 R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan
 Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
 A:Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endo
 A:Reference number: A95842; MUID:21396508; PMID:11481431
 A:Accession: E95934
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-307 <KUR>
 A:Cross-references: UNIPROT:Q92VG4; GB:AL591985; PIDN:CAC49141.1; PID:gl51140626; GSPDB:C
 A:Experimental source: strain 1021, megaplasmid pSymB
 R:Galibert, F.; Finan, T.M.; Long, S.R.; Fuhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
 Pella, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
 L.; Hyman, R.W.; Jones, T.
 Science 293, 668-672, 2001
 A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
 hebaault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
 A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
 A:Reference number: A96039; MUID:21368234; PMID:11474104
 A:Contents: annotation
 C:Genetics:
 A:Gene: SM521237
 A:Genome: plasmid

Query Match 5.2%; Score 107.5; DB 2; Length 307;
 Best Local Similarity 20.6%; Pred. No. 0.064;
 Matches 78; Conservative 46; Mismatches 131; Indels 123; Gaps 17;

Qy 32 PSSPAGGEDRVHLVLSW-RSGSFLGQLFSQHPDV- -YLMPEAPWHVMTTILSQGSA 87
 Db 2 PSQP- - - - -VRIAYIAGVGRSGTILDIALGQAAVVGAGEITSLTRHV- - -RHNEYC 52
 Qy 88 TLHMAVRLMRSIFLCD- - - - -MDVFADYMPQSRNLSPFNWATSRAL 130
 Db 53 ACQNAIRD- - - - -CSFWSSVRREWSGDQDPLMEYCALQQKFEGLSMTRLGSGMGL 105
 Qy 131 CSPPACSAFPRGTISKQDVCKTLCTROPFSLAREACRSYSHVVLKEVFFNQLVLYPLLS 190
 Db 106 - - - - -GKQFSYLILHTKRLFS-AMQSC-SGRQIVDVSSKLPGRAMAVAQI- 148
 Qy 191 DPALMLRLVHLVRDPAVRLSREAGPILARDNGIVLGTNGKWKVBADPHLRILIREVCRSH 250
 Db 149 -PGIDMRVHLVRDGRGV- - - - -AWSLL- - - - -KGYERDAKSGLOKEI- - - 185
 Qy 251 VRIABAATLKPPPLRGYR- - - - -LVRFEDLAREPLAERIALY 289
 Db 186 - - - - -KPKSVFPTALRSMWNLAVEYLSRKLGSKVMRVRYEDFASDPVAVMOQIG 236
 Qy 290 AFTGLTLTPLEAWTHNTHGSGICKPTEAFHTSSRN- - - - -ARNVSOAWRHLPPT 341
 Db 237 TFLDLDS-QVGTSEN- - - - -GEAMGPHQHVAGNRLRMWASIALNKDETWRTRMPAR 288

Qy 342 KILRVQEVACAGALQLGY 359
 Db 289 QQVVSFQRLGGWMLRYGY 306

RESULT 6
 T16350
 hypothetical protein F42G9.8 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
 C:Accession: T16350
 R:Taich, A.
 submitted to the EMBL Data Library, March 1996
 A:Description: The sequence of C. elegans cosmid F42G9.
 A:Reference number: Z18498
 A:Accession: T16350
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-359 <TAI>
 A:Cross-references: UNIPROT:Q20351; EMBL:U00051; NID:gl216308; PIDN:AAA913.
 A:Experimental source: strain Bristol N2
 C:Genetics:
 A:Gene: CESP:F42G9.8
 A:Introns: 28/2; 50/2; 113/2; 155/3; 220/3; 290/2
 C:Superfamily: Caenorhabditis elegans hypothetical protein F42G9.8

Query Match 5.1%; Score 105; DB 2; Length 359;
 Best Local Similarity 19.8%; Pred. No. 0.13;
 Matches 80; Conservative 53; Mismatches 131; Indels 140; Gaps 21;

Qy 11 VTVLLLAQTTCLLLF- - - - -ISRPGSPSPAGGEDRVH- - - - -VLV 46
 Db 31 IYIFIFCTICLLIFSSIKCKLQKLEQLSKESLIFNQDARHSRRLSNLEQLIFV 90
 Qy 47 LSSWRSGSFLGQLFSQHPDV- - - - -FYLMPEAPWHVMTTILSQGSAATLHMAVR 94
 Db 91 GGVPRSGTTLRAILDADHPDRCGCTGTMLLPSFLTQAGWENDV- - -NNSGIT- - - 141
 Qy 95 DLMSRIFLCMDVDYAPMYSQSRNLSPFNWATSRALCSPPACSAFPRGTISK-QDVCKTL 153
 Db 142 - - - - -QEVDP- - - - -DAVSAPITEIVAKHSELAPRL 167
 Qy 154 CTROPFSLAREACRSYSHVVLKEVFPFNQLVLYPLLSDPALNLRVHLPDPAVRLSR- 212
 Db 168 CNKDP- - - - -YTALMLPTIR- - - - -RLYP- - - - -NAKFLMIRDARAVVHSMI 205
 Qy 213 EAAGPILARDNGIVLGTNGKWKVBADPHLRILIREVCRSHVRIAEATLKPPPLRGYRLV 272
 Db 206 ERKVPVAGYNTSDISMPQW- - - - -NQELRKMTFCNN- - - - -APGQCIK- - - - -V 247
 Qy 273 RFEDLAREPLAERIALYAPTGLTLTPLEAWTHNTHGSGICKPI- - - - -EAPHTSS-RNA 327
 Db 248 YVERLIQKPAEILRITNFDLPFSQW- - - - -LRHQDLICDEVLDNDQSFASQVKN 301
 Qy 328 RNVS- - - - -QAWRHLPPTKILRVQEVACAGALQLGY- - - - -RPVYS 364
 Db 302 INKALTSTWDFCFSEETLRKLDV-APPLGLTGLYDTSISKPDYS 344

RESULT 7
 G70729
 hypothetical protein Rv2267c - Mycobacterium tuberculosis (strain H37Rv)
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Jul-2004
 C:Accession: G70729
 R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.;
 Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A:Reference number: A70500; MUID:98295987; PMID:9634230
 A:Accession: G70729
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-388 <COL>
 A:Cross-references: UNIPROT:Q50695; GB:Z77163; NID:gl216308; PIDN:CAB00968.1
 A:Experimental source: strain H37Rv
 C:Genetics:
 A:Gene: Rv2267c

Db 128 NIIMTGNSTCLDIALRMLNTRGDSILVEKYSFSPALQSMRPLGLSCIPIDMDQF-GFLP 186
 Qy 114 QSRNLSAFFNW-ATSRALCSPACSPAPRGITISQDVCKTCTRQPFSLAREACRSYSHV 172
 Db 187 ESMD-DILTNDWATSYGSPKPHVLYTPTGQNPSTLSVERRKQIYTLAQK-----HDII 241
 Qy 173 VLKVRFNQV-LY-----PLISDPAALMLRIVHLVRDPAVLRSLREAAGPILARN-----GI 225
 Db 242 ILEDEPVYQLMDAYEGKPEAAKRAFTNE--QFKELIPSLSDMDVGRVIRMDLSKVV 299
 Qy 226 VLGTNGKWEADPHRLRILREVCRSHVRIAEATLKPFPFLRG-RYRLVRP--EDLAREPL 282
 Db 300 APGSRVGNFTAQPLF-----IERGLAAETATQASISQGIYAMPKHWGQGYLEWL 353
 Qy 283 AEIRALYAFT 292
 Db 354 KHIR--YSYT 361
 RESULT 11
 S78246
 N:Altepidase Clp (EC 3.4.21.-) ATP-binding chain clpC [similarity] - Odontella sinensis
 N:Altepidase names: ATP-dependent Clp proteinase regulatory chain; caseinolytic Clp prote
 N:Contains: adenosinetriphosphatase (EC 3.6.1.3)
 C:Species: chloroplast Odontella sinensis
 C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 09-Jul-2004
 C:Accession: S78246
 P:Kowalik, K.V.; Stoebe, B.; Schaffran, I.; Kroth-Pancic, P.; Freier, U.
 Plant Mol. Biol. Rep. 13, 336-342, 1995
 A:Title: The Chloroplast Genome of a chlorophyll a+c containing Alga, Odontella sinensis
 A:Reference number: S78238
 A:Accession: S78246
 A:Molecule type: DNA
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Residues: 1-885 <KOW>
 A:Cross-references: UNIPROT:P49574; EMBL:Z67753; NID:g1185127; PIDN:CAA91619.1; PID:g118
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1995
 C:Genetics:
 A:Gene: clpC
 A:Genome: chloroplast
 C:Function:
 A:Description: allows clpP to hydrolyze polypeptides and proteins, probably by a chaperon
 e activity; ATP hydrolysis is required for Clp binding of proteins but not of smaller
 C:Superfamily: endopeptidase Clp ATP-binding chain
 C:Keywords: ATP; chloroplast; hydrolase; molecular chaperone; nucleotide binding; P-loop
 F:218-225/Region: nucleotide-binding motif A (P-loop)
 F:285-290/Region: nucleotide-binding motif B
 F:560-567/Region: nucleotide-binding motif A (P-loop)
 F:628-633/Region: nucleotide-binding motif B
 F:224/Binding site: ATP (Lys) #status predicted
 F:566/Binding site: ATP (Lys) #status predicted
 Query Match 4.6%; Score 95; DB 1; Length 885;
 Best Local Similarity 21.0%; Pred. NO. 3.5;
 Matches 89; Conservative 59; Mismatches 161; Indels 114; Gaps 18;
 Qy 5 RPSKTVTVLLAQ-----TTCALFII-SRPGSSPAGGEDRVHVLVLSWR 51
 Db 4 KFTGAIKVIMLSQSEARMGNHFNVTGTEQLGIIQGRHGIGARALKQKV---TLKKAR 60
 Qy 52 -----SGSFLGQLPSQHPDVFYLMPEAWHVTTLSSQSAATLHMAVRDLMS--IF 101
 Db 61 REIELYIGRGTFVASEIPPTPRAKRVLEMAVHEGKDLQGNFVGTETHEILLALISESDGVA 120
 Qy 102 LCDMDVDPAYPQSRNLSAFFNWATSRALCSPACSPAPRGITISQDVCKTCTRQPFSL 161
 Db 121 MRTLKLGWNPVKLRNLMLVI-----EENQEEILRPLTQAEKFL 161
 Qy 162 AREACRS-----YSHVVLKVRFNQVLYPLISDPAALML-RIVHLVRDPAVLRSLRE 213
 Db 162 EREKKGSTPTLDEYSENISEKAVDGLK-----DPVIGRDKEIHEV---IKVLARR 210
 Qy 214 AAGPILARDNGIVLGTNGKWEADPHRLRILREVCRSHVRIAEATLKPFPFLRGYRLVR 273

Db 211 KNPVLIGEPGV-----GKTAVAEGLAQLI-----IAE-----KAPFLDGNL----- 248
 Qy 274 FEDLAREPLAEIRALYAFTGLTLTPQLEAWIHNITHSGIGKPIEAFHT-----SSRNAR 328
 Db 249 ---LMALDGLSILAGTKYRG-EFEERIKRIVEEVQNDSSAILVIDEITLTVGAGAEGAV 304
 Qy 329 NVSQARWALPFTKILRVQVVCAGALQLLYR-----PVYSAQQOQDUTLTD 374
 Db 305 DAANILKPALARGKFR-----CIGATTIDEYKRIERDPALERRFPVHVKEPTVGVTIE 359
 Qy 375 LVL 377
 Db 360 ILL 362
 RESULT 12
 VGMJBV
 peplomer glycoprotein precursor - Berne virus (strain P138/72)
 C:Species: Berne virus
 C:Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 09-Jul-2004
 C:Accession: A36759
 P:Snijder, E.J.; Den Boon, J.A.; Spaan, W.J.M.; Weiss, M.; Horzinek, M.C.
 Virology 178, 355-363, 1990
 A:Title: Primary structure and post-translational processing of the Berne virus peplomer
 A:Reference number: A36759; MUID:91020973; PMID:2219698
 A:Accession: A36759
 A:Molecule type: genomic RNA
 A:Residues: 1-1581 <SNI>
 A:Cross-references: UNIPROT:P23052; GB:X52506; NID:g62059; PIDN:CAA36748.1; PID:g62060
 C:Genetics:
 A:Gene: P
 C:Superfamily: Berne virus peplomer glycoprotein
 C:Keywords: glycoprotein; transmembrane protein
 F:1-19/Domain: signal sequence #status predicted <SIG>
 F:20-1581/Product: peplomer glycoprotein #status predicted <PGP>
 F:1547-1572/Domain: transmembrane #status predicted <TMN>
 F:25,310,384,494,574,935,969,1267,1297,1327,1385,1389,1428,1431,1438,1483,1487,1495,1515
 Query Match 4.6%; Score 94.5; DB 1; Length 1581;
 Best Local Similarity 17.6%; Pred. No. 8.3;
 Matches 81; Conservative 65; Mismatches 135; Indels 179; Gaps 19;
 Qy 2 WLPFRFSKTVTVLLAQTTCLLPIISRPSPAGGEDRVHVLVLSWRSGSFLGQLF 61
 Db 259 WYVAFQNKATAVILPSELIVPAQKVT-----RLG 289
 Qy 62 SQHPDVFYLMPEAWHVTTLSSQSAATLHMAVRDLMSIFLCDMDVFDAYMPQSRNLSAF 121
 Db 290 VNTPDYFVLVKQAVH-----YLSQA-NLSP- 313
 Qy 122 FHWATSRALCSPACSPAPRGITISQDVCKTCTRQPFSLAREACRSYSHV-LKEVRFF 180
 Db 314 -NYALFSALCN-----SLYQOQATSLTLCGSPFFVAQEC---YNNALYLPDAVFT 360
 Qy 181 NL-----QVLYPL-----LSDPALML-----RIVHLVRDPAVL----- 209
 Db 361 TLESTLFSWDYQINFLNOVLTONETFLQLPATNYQGTLSQGRMLNLFDAIVFLDFFD 420
 Qy 210 -----RSREAGP-----ILARDNGIVLGTNGKWEADPHRLRILREVCRSHVRIAEATLKP 261
 Db 421 TKFVRTNDAPSSDIFVYVVARQAQLIRYGNFRIEQINGYFQV---KCSSNI-----ISTLEP 473
 Qy 262 PPLRGYRLVRFPDLAREPLAEIRALYAFT----- 292
 Db 474 HP--AGVIMTARHSHMWSVAARNSTSPYCVTHSLTTFGKLDISTSWFFHTLALSPGVSQ 531
 Qy 293 -----GLTLPQLEAWIHNIT-----HGSGIGKPIEAFHTSSRNARVS 331
 Db 532 VSMPLLSAAVGVMYHPIEHPIPLLLLAQSQYQPSFFNIGINKTI-TLTTLOQAYAVY 590
 Qy 332 QAWRHALPFTKILRVQVVCAGALQLLYRVPVYSADQQDRL 371

Db 591 TAMFLSVYVRLPEARLTLG-VOLVFFIOALLSIKQADL 629

RESULT 13

A70841

probable oxalyl-CoA decarboxylase - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 16-Aug-2004

C:Accession: A70841

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998

A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: A70841

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-582 <COL>

A:Cross-references: UNIPROT:O53639; GB:AL021926; GB:AL123456; NID:g3261520; PIDN:CAA1731

A:Experimental source: strain H37Rv

C:Genetics:

A:Gene: oxaA

C:Superfamily: Acetolactate synthase, large subunit/pyruvate oxidase; thiamin pyrophosph

Query Match 4.6%; Score 94; DB 2; Length 582;

Best Local Similarity 21.4%; Pred.No.2.5; Mismatches 44; Indels 60; Gaps 11;

Matches 55; Conservative 44;

QY 152 TCTROPFSLAREACRSYSHVVLKEVFFNQLVLPILLSDP-----ALNLRIVHLV 202

DB 2 TTRSASCTVLTGDC-----HLVDALKANDVITYGVGIPITDLARAAQASGIRYIGF- 56

QY 203 RDPFAVLUREAGPILARONGIVLTNGK-WFADPHLRLIREVCKSHVRIAEATLKP 261

DB 57 RHEASAGNAAAAGFLTARP-GVCLTTSGPGLNGLPALANATTNCPMTQISGSSS-RP 114

QY 262 -----PPFLGRYLVRFEDLAREPLAEIRALYAFTGLTLPQLEAW 303

DB 115 MVDLQRGDYDQLNNAARFVKAAYRIGQVDIGRGVARAIR-----TATSGRPGG 166

QY 304 IHNITHGSGIKPTAEPTSSRNARNVSOQRHALPFTKILRVQEVQAGALQLLGYRPVY 363

DB 167 VYLDIPGDLGQAVEASAAS-----GAIWRPVDPAPELLPAPEALDRALDVL----- 213

QY 364 SAQQQRDLTLDLVLPRG 380

DB 214 -AQQRPL---LVLSKG 226

RESULT 14

C75278

DNA polymerase III, tau/gamma subunit - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004

C:Accession: C75278

R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma S.; Smith, H.O.; Venter, J.C.; Fraser, C.M. Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A:Reference number: A75250; MUID:20036896; PMID:10567266

A:Accession: C75278

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-615 <WHI>

A:Cross-references: UNIPROT:Q9RRS5; GB:AE002071; GB:AE000513; NID:g6460218; PIDN:AAF1195

A:Experimental source: strain R1

C:Genetics:

A:Gene: DR2410

A:Map position: 1

C:Superfamily: DNA-directed DNA polymerase III gamma chain

Query Match 4.6%; Score 93.5; DB 2; Length 615;

Best Local Similarity 23.4%; Pred.No.2.9;

Matches 92; Conservative 41; Mismatches 141; Indels 119; Gaps 21;

QY 4 PRPSSKTVTVLLAQTTCULLLFIISRPSPSPAGGEDRVHVLVLSWSRSGSSFLGQLFSQ 63

DB 44 PRGVGKTTTARLIAMTA-----NCTGPAPKPCGCECLAV-----RAGS----- 83

QY 64 HDPVFLMEPAWHVWTLTSGSNAATHMA--VRDLMRISIFLCDDM-----VFDAYMQPS 115

DB 84 HPDVMIEI-----DAASNNVDVDRLEKVGKLAAMRGKKIYIIDEAHMMS 129

QY 116 RNLSAFENWATSRALCSPSPACSAFPRGTITSKDVCKTCTROPFSLAREACRSY--SHVV 173

DB 130 R--AAFN-ALLKTLSEPEHVFILATTEPEKIIFILSR-----CQHYRFRRLT 176

QY 174 LKEV--RFFNLQVLPILLSDP-ALNL--RIVH-LVRDPRAVLRSRAAGPILARDNGIVL 227

DB 177 SEEIACKLAGLVLTLEGASADPDALNLIGRLADGAMRDGESSLLERMLAAGTAVTRP----- 231

QY 228 GTNGKWEADPHLRLIREVCRSHVRIAEATLKPFPFLGRYLVRFEDLAREPLAEIRA 287

DB 232 -----AVEEALGUPPGERVGRVASALLVGD--AGEAISGAQA 266

QY 288 LY--AFTGLTLTPQLEAWIHNITHGS-GIGKPIEAFHTSSRNARNVSOQRHALPFTKIL 344

DB 267 LYRDGFAARTVVEGLVAFAFGAALHAELGLGE-----EGLEGAEVPRLL 310

QY 345 RVQEVQAGALQLLGYRPVYSADQQRDLTLDVL 377

DB 311 KLQ----AALDEQEARFARSADQQ---SLELAL 336

RESULT 15

B83629

probable ATP-binding component of ABC transporter PA0136 [imported] - Pseudomonas aerugin

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004

C:Accession: B83629

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bri adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, ; Lory, S.; Olson, M.V. Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathox

A:Reference number: A82950; MUID:20437337; PMID:10984043

A:Accession: B83629

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-523 <STO>

A:Cross-references: UNIPROT:Q91626; GB:AE004451; GB:AE004598; NID:g9945958; PIDN:AAG03526

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA0136

Query Match 4.5%; Score 93; DB 2; Length 523;

Best Local Similarity 26.4%; Pred.No.2.6;

Matches 37; Conservative 17; Mismatches 60; Indels 26; Gaps 5;

QY 13 VLLLAQTTCULLLFIISRPSPSPAGGEDRVHVLVLSWSRSGSSFLGQLFSQHPDVFYLM 72

DB 384 VLALAE-EIIRRFAPKAPGAGAPARS-----LSGGNLQKFILGREILQAPRLVAAH 434

QY 73 PAWHVWTLTSGSNAATH--NAVRDLMRISIFLCDDVDVDAYMQPSRNLSAFPNWATSR 129

DB 435 PTMGV----DVGAALIHRLALRDAGTAVLVVSEDLDELFLLSDR-----IAA 480

QY 130 LCSPAPACSAFPRGTITSKDV 149

DB 481 LCSGRCLCPAVATASAPQKV 500

Search completed: June 23, 2005, 08:50:12

Job time : 9.8459 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2005, 08:33:13 ; Search time 43.3472 Seconds
(without alignments)
4666.302 Million cell updates/sec

Title: US-10-697-828-7

Perfect score: 2057

Sequence: 1 MRLPRFSSTVMSLLMVQTG.....LPRGMDSPKWSSTKQPES 395

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

UniProt_03.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|--------------|------------------------|
| 1 | 2057 | 100.0 | 395 | 2 Q9QUP4 | Q9QUP4 mus musculus |
| 2 | 1708 | 83.0 | 395 | 2 Q9GZK3 | Q9GZK3 homo sapien |
| 3 | 1536.5 | 74.7 | 390 | 2 Q7LCH3 | Q7LCH3 homo sapien |
| 4 | 1536.5 | 74.7 | 411 | 2 Q9GZS9 | Q9GZS9 homo sapien |
| 5 | 1187 | 57.7 | 392 | 2 Q6GN39 | Q6GN39 xenopus lae |
| 6 | 1022.5 | 49.7 | 386 | 2 Q9NCOS | Q9NCOS homo sapien |
| 7 | 1017.5 | 49.5 | 386 | 2 Q9Y5R3 | Q9Y5R3 homo sapien |
| 8 | 1010.5 | 49.1 | 370 | 2 Q9IV46 | Q9IV46 homo sapien |
| 9 | 991 | 48.2 | 388 | 2 Q9R111 | Q9R111 mus musculus |
| 10 | 984 | 47.8 | 388 | 2 Q9WUE5 | Q9WUE5 mus musculus |
| 11 | 640 | 31.1 | 483 | 2 Q794G9 | Q794G9 mus musculus |
| 12 | 640 | 31.1 | 530 | 2 Q88276 | Q88276 mus musculus |
| 13 | 637 | 31.0 | 530 | 2 Q90WV3 | Q90WV3 mus musculus |
| 14 | 636.5 | 30.9 | 483 | 2 Q9UED5 | Q9UED5 homo sapien |
| 15 | 636.5 | 30.9 | 530 | 2 Q9Y4C5 | Q9Y4C5 homo sapien |
| 16 | 585 | 28.4 | 420 | 2 Q6DBY9 | Q6DBY9 brachydanio |
| 17 | 581 | 28.2 | 484 | 2 Q9EP78 | Q9EP78 m n-acetylgl |
| 18 | 580 | 28.2 | 486 | 2 Q75667 | Q75667 homo sapien |
| 19 | 579 | 28.1 | 486 | 2 Q9NS84 | Q9NS84 homo sapien |
| 20 | 578 | 28.1 | 484 | 2 Q9N8B0 | Q9N8B0 mus musculus |
| 21 | 576 | 28.0 | 485 | 2 Q9XQ88 | Q9XQ88 rattus norv |
| 22 | 563.5 | 27.4 | 479 | 2 Q7LGC8 | Q7LGC8 homo sapien |
| 23 | 561.5 | 27.3 | 479 | 2 Q75099 | Q75099 mus musculus |
| 24 | 551 | 26.8 | 411 | 2 Q9EQC0 | Q9EQC0 mus musculus |
| 25 | 550 | 26.7 | 472 | 2 Q88199 | Q88199 mus musculus |
| 26 | 549 | 26.7 | 411 | 2 Q43916 | Q43916 homo sapien |
| 27 | 542 | 26.3 | 474 | 2 Q9QZL2 | Q9QZL2 rattus norv |
| 28 | 495 | 24.1 | 458 | 1 C6ST_CHICK | C6ST_CHICK gallus gall |
| 29 | 457.5 | 22.2 | 441 | 2 Q93403 | Q93403 torpedo cal |
| 30 | 373.5 | 18.2 | 257 | 2 Q79415 | Q79415 mus musculus |
| 31 | 320.5 | 15.6 | 225 | 2 Q6RY62 | Q6RY62 cavia porce |

RESULT 1

| | | | |
|--------|---|--|---------|
| Q9QUP4 | PRELIMINARY; | PRT; | 395 AA. |
| ID | Q9QUP4 | | |
| AC | Q9QUP4; | | |
| DT | 01-MAY-2000 (TrEMBLrel. 13, Created) | | |
| DT | 01-MAY-2000 (TrEMBLrel. 13, Last sequence update) | | |
| DT | 05-JUL-2004 (TrEMBLrel. 27, Last annotation update) | | |
| DE | N-acetylglucosamine 6-O-sulfotransferase. | | |
| GN | Name=Chat5; Synonyms=I-GlcNAc-6-ST; | | |
| OS | Mus musculus (Mouse). | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | |
| OC | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. | | |
| OX | NCBI_TaxID=10090; | | |
| RN | [1] | | |
| RP | SEQUENCE FROM N.A. | | |
| RC | STRAIN=C57BL/6; TISSUE=Intestine; | | |
| RC | MEDLINE=99423499; PubMed=10491328; DOI=10.1006/bbrc.1999.1324; | | |
| RA | Lee J.-K., Bhakta S., Rosen S.D., Hemmerich S.; | | |
| RT | "Cloning and characterization of a mammalian N-acetylglucosamine-6- | | |
| RT | sulfotransferase that is highly restricted to intestinal tissue."; | | |
| RL | Biochem. Biophys. Res. Commun. 263:543-549(1999). | | |
| RN | [2] | | |
| RP | SEQUENCE FROM N.A. | | |
| RC | STRAIN=C57BL/6; TISSUE=Intestine; | | |
| RA | Lee J.-K., Bhakta S., Rosen S.D., Hemmerich S.; | | |
| RL | Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases. | | |
| DR | EMBL; AF176841; AAD56003.1; - | | |
| DR | EMBL; AF176840; AAD56002.1; - | | |
| DR | MGD; MGI:1931825; Chat5. | | |
| DR | GO; GO:0005615; C:extracellular space; TAS. | | |
| DR | InterPro; IPR000863; Sulfotransferase. | | |
| DR | Pfam; PF00685; Sulfotransfer_1; 1. | | |
| KW | Transferase. | | |
| SQ | SEQUENCE 395 AA; 44537 MW; 3FDF71E43ED383BE CRC64; | | |
| | Query Match 100.0%; Score 2057; DB 2; Length 395; | | |
| | Best Local Similarity 100.0%; Pred. No. 1.4e-168; | | |
| | Matches 395; Conservative 0; Mismatches 0; Indels 0; Gaps 0; | | |
| Qy | 1 | MRLPRFSSTVMSLLMVQTGILVFLVSRQVPSSPAGLGERVHVLVLSWSRSGSFGOLF | 60 |
| Db | 1 | MRLPRFSSTVMSLLMVQTGILVFLVSRQVPSSPAGLGERVHVLVLSWSRSGSFGOLF | 60 |
| Qy | 61 | SOHPDVFYLMPEPAWHVMDTLRSQGSAPALHMAVRDLIRSVFLCDMDVFDAYLPWRNITSDL | 120 |
| Db | 61 | SOHPDVFYLMPEPAWHVMDTLRSQGSAPALHMAVRDLIRSVFLCDMDVFDAYLPWRNITSDL | 120 |
| Qy | 121 | FOAVSRALCSPPVCEAFARGNISSEVCKPLCATRPFGLAQEACSSYSHVVLKEVRFN | 180 |
| Db | 121 | FOAVSRALCSPPVCEAFARGNISSEVCKPLCATRPFGLAQEACSSYSHVVLKEVRFN | 180 |
| Qy | 181 | LQVLYPLLSPPALNLRIVHLVRDPAVLRREQTAKALARDNGIVLTNGTWTWEADPRLR | 240 |
| Db | 181 | LQVLYPLLSPPALNLRIVHLVRDPAVLRREQTAKALARDNGIVLTNGTWTWEADPRLR | 240 |

Q7Q1F9 anopheles g
Q677T4 lymphocyeti
Q9VMC3 drosophila
Q9VMC4 drosophila
Q95JA8 oryctolagus
Q9AD05 mus musculus
Q8MZD1 drosophila
Q92V94 thriobium m
Q93J66 streptomyce
Q6ZET4 synechocyst
Q8PRA0 xanthomonas
P64963 mycobacteri
P64964 mycobacteri
Q68RS5 prochloron

```
QY 241 VVNEVCSSHVRIAEAAHKKPPFLQDRYRLRVYEDLARDPLTVIRELYAFTGLTPQLQ 300
Db 241 VVNEVCSSHVRIAEAAHKKPPFLQDRYRLRVYEDLARDPLTVIRELYAFTGLTPQLQ 300
QY 301 TWIHNTHGSGPGARREAFKTTSRDALSVSQAWRHHTLPFAKIRRVQELCGALQLLGYS 360
Db 301 TWIHNTHGSGPGARREAFKTTSRDALSVSQAWRHHTLPFAKIRRVQELCGALQLLGYS 360
QY 361 VVSELEQDRLSLDLLPRGMDSPKWSSTKQEPES 395
Db 361 VVSELEQDRLSLDLLPRGMDSPKWSSTKQEPES 395

RESULT 2
Q9GZX3 Q9GZX3 PRELIMINARY; PRT; 395 AA.
AC Q9GZX3;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE N-acetylglucosamine 6-O-sulfotransferase GST-4beta (Corneal N-
DE acetylglucosamine-6-O-sulfotransferase) (Carbohydrate (N-
DE acetylglucosamine 6-O) sulfotransferase 6).
GN Name=GST4beta; Synonym=CHST6;
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21096027; PubMed=1181564; DOI=10.1093/glycob/11.1.75;
RA Hemmerich S., Lee J.-K., Bhakta S., Bistrup A., Ruddle N.R.,
RA Rosen S.D.;
RA "Chromosomal localization and genomic organization of the galactose/N-
RA acetylglucosamine/N-acetylglucosamine 6-O-sulfotransferase gene
RA family."
RL Glycobiology 11:75-87(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20472330; PubMed=11017086; DOI=10.1038/79987;
RA Akama T.O., Nishida K., Nakayama J., Watanabe H., Ozaki K.,
RA Nakamura T., Dota A., Kawasaki S., Inoue Y., Maeda N., Yamamoto S.,
RA Fujiwara T., Thonar E.J., Shimomura Y., Kinoshita S., Tanigami A.,
RA Fukuda M.N.;
RA "Macular corneal dystrophy type I and type II are caused by distinct
RA mutations in a new sulphotransferase gene."
RL Nat. Genet. 26:237-241(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L.H., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,
RA Dratchenko L., Marudina K., Farmer A.A., Rubin G.M., Hong L.,
RA Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Heiton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzyzanski M.I., Skalska U., Smalus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RA "Generation and initial analysis of more than 15,000 full-length human
RA and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [4]
```

```
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RA Director MGC Project;
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF280086; AAG48244.1; -
DR EMBL; AF219991; AAG26327.1; -
DR EMBL; BC074883; AAH74883.1; -
DR EMBL; BC074834; AAH74834.1; -
DR EMBL; AF219990; AAG26325.1; -
DR Genew; HGNC:6938; CHST6.
DR GO; GO:0005794; C:Golgi apparatus; TAS.
DR GO; GO:0005157; F:N-acetylglucosamine 6-O-sulfotransferase ac. .; TAS.
DR GO; GO:0060444; P:N-acetylglucosamine metabolism; IC.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
KW Transferase.
SQ SEQUENCE 395 AA; 44098 MW; 433CA60248A48F67 CRC64;

Query Match 83.0%; Score 1708; DB 2; Length 395;
Best Local Similarity 83.0%; Pred. No. 1.6e-138;
Matches 328; Conservative 24; Mismatches 43; Indels 0; Gaps 0;

QY 1 MRLPRFSSTWMLSLMVQGTGILVLSRVQSPSPAGLGRVHVLVLSWRSWSSSFVQQLF 60
Db 1 MWLPRVSVTAVTALLLAQFTLLFLVSRPCSPSPAGGEARVHVLVLSWRSWSSSFVQQLF 60

QY 61 SQHPDVFYLMPEPAWHVWDTLSQGSAPALHMAVDRDLIRSVFLCDMDVDFDAYLPWRRNLSDL 120
Db 61 NQHPDVFYLMPEPAWHVWDTLSQGSAAATLHMAVDRDLVRSVFLCDMDVDFDAYLPWRRNLSDL 120

QY 121 FQWVSVRALCSPPVCEAFARGNISSEEVCKPKCATRPFGLAOEACSSYSHVVLKEYRFFN 180
Db 121 FQWVSVRALCSPPVCEAFARGNISSEEVCKPKCATRPFGLAOEACSSYSHVVLKEYRFFN 180

QY 181 LQVLYPLLSDDPALNLRIVHLRDPRAVLSRQTKALARDNGIVLGTGTWVEADPGLR 240
Db 181 LQVLYPLLSDDPALNLRIVHLRDPRAVLSRQTKALARDNGIVLGTGTWVEADPGLR 240

QY 241 VVNEVCSSHVRIAEAAHKKPPFLQDRYRLRVYEDLARDPLTVIRELYAFTGLTPQLQ 300
Db 241 VVNEVCSSHVRIAEAAHKKPPFLQDRYRLRVYEDLARDPLTVIRELYAFTGLTPQLQ 300

QY 301 TWIHNTHGSGPGARREAFKTTSRDALSVSQAWRHHTLPFAKIRRVQELCGALQLLGYS 360
Db 301 TWIHNTHGSGPGARREAFKTTSRDALSVSQAWRHHTLPFAKIRRVQELCGALQLLGYS 360

QY 361 VVSELEQDRLSLDLLPRGMDSPKWSSTKQEPES 395
Db 361 VVSELEQDRLSLDLLPRGMDSPKWSSTKQEPES 395

RESULT 3
Q7LCN3 Q7LCN3 PRELIMINARY; PRT; 390 AA.
AC Q7LCN3;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE N-acetylglucosamine 6-O-sulfotransferase.
GN Name=I-GlcNAC-6-ST;
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99423499; PubMed=10491328; DOI=10.1006/bbrc.1999.1324;
RA Lee J.K., Bhakta S., Rosen S.D., Hemmerich S.;
RA "Cloning and characterization of a mammalian N-acetylglucosamine-6-
RA sulfotransferase that is highly restricted to intestinal tissue."
RL Biochem. Biophys. Res. Commun. 263:543-549(1999).
RN [2]
```

RP SEQUENCE FROM N.A.
RC TISSUE=Intestine;
RA Lee J.-K., Bhakta S., Rosen S.D., Hemmerich S.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF176839; AAD56001.1; -;
DR EMBL; AF176838; AAD56000.1; -;
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; P:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransferase.
KW Transferase.
SQ SEQUENCE 390 AA; 43832 MW; 132D1C3B6BE38179 CRC64;

Query Match 74.7%; Score 1536.5; DB 2; Length 390;
Best Local Similarity 76.0%; Pred. No. 9.4e-124;
Matches 295; Conservative 30; Mismatches 62; Indels 1; Gaps 1;

Qy 1 MRLPRFSSTVMSLLMVQTC-ILVFLVSRQVPPSPAGLGERVHVLLVSSWRSGSFGQL 59
Db 1 MMLPRFSSTVTVLLLAQTTCLLLFIIISRPSPAGGEDRVHVLLVSSWRSGSFGQL 60

Qy 60 FSOHPDVFLMEPAWHVWDTLSQSPALHMAVRDLIRSVFLCDMDVDFDAYLPWRRNISD 119
Db 61 FSOHPDVFLMEPAWHVWDTLSQSPALHMAVRDLIRSVFLCDMDVDFDAYLPWRRNISD 120

Qy 120 LFQWAVSRALCSPPVCEAFARGNISSEEVCKPLCATRPFGLAQACSSYSHVWLKEVRFF 179
Db 121 FFWATSRALCSPPACSAFPRTGTSKQDVCKTLCTROPFSLAREACRSYSHVWLKEVRFF 180

Qy 180 NLQVLYPLLSDDPALNLRIVHLVRDPRVAVLSRQVATKALARDNGIVLGTNGKWEADPHL 239
Db 181 NLQVLYPLLSDDPALNLRIVHLVRDPRVAVLSRQVATKALARDNGIVLGTNGKWEADPHL 240

Qy 240 RVNVEVCRSHVRIAEALHKPPFPLODRYRLVRYEDLARDPLTVIRELYAFTGLTLPOL 299
Db 241 RLIREVCRSHVRIAEALHKPPFPLODRYRLVRYEDLARDPLTVIRELYAFTGLTLPOL 300

Qy 300 QTWHNTHGSGPGARREAFKTTSRDALSVSQAWRHLPFAKIRRVQELCGALQLGYR 359
Db 301 EAWHNTHTGSGIGKPIEAFTSSRNARNVSVQAWRHLPFAKIRRVQELCGALQLGYR 360

Qy 360 SVHSELEQRDLSDLLPRGMDSPKWS 387
Db 361 PVYSADQQRDLTLDVLRPGDPHFWSAS 388

RESULT 4
ID Q9GZS9 PRELIMINARY; PRT; 411 AA.
AC Q9GZS9; Q9UBV3;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)
DE Intestinal GlcNAc-6-sulfotransferase (Intestinal N-acetylglucosamine-6-O-sulfotransferase).
GN Name=CHST5;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
ON NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20472330; PubMed=11017086; DOI=10.1038/79987;
RA Akana T.O., Nishida K., Nakayama J., Watanabe H., Ozaki K.,
RA Nakamura T., Dotsa A., Kawasaki S., Inoue Y., Maeda N., Yamamoto S.,
RA Fujiwara T., Thonar E.J., Shimomura Y., Kinoshita S., Tanigami A.,
RA Fukuda M.N.;
RT "Macular corneal dystrophy type I and type II are caused by distinct mutations in a new sulphotransferase gene."
RL Nat. Genet. 26:237-241(2000).
DR EMBL; AF246718; AAG28023.1; -;
DR EMBL; AF219991; AAG26326.1; -;
DR Genew; HGNC:1973; CHST5.

DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; P:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransferase.
KW Transferase.
SQ SEQUENCE 411 AA; 46160 MW; 97642D54BE926E06 CRC64;

Query Match 74.7%; Score 1536.5; DB 2; Length 411;
Best Local Similarity 76.0%; Pred. No. 1e-123;
Matches 295; Conservative 30; Mismatches 62; Indels 1; Gaps 1;

Qy 1 MRLPRFSSTVMSLLMVQTC-ILVFLVSRQVPPSPAGLGERVHVLLVSSWRSGSFGQL 59
Db 22 MMLPRFSSTVTVLLLAQTTCLLLFIIISRPSPAGGEDRVHVLLVSSWRSGSFGQL 81

Qy 60 FSOHPDVFLMEPAWHVWDTLSQSPALHMAVRDLIRSVFLCDMDVDFDAYLPWRRNISD 119
Db 82 FSOHPDVFLMEPAWHVWDTLSQSPALHMAVRDLIRSVFLCDMDVDFDAYLPWRRNISD 141

Qy 120 LFQWAVSRALCSPPVCEAFARGNISSEEVCKPLCATRPFGLAQACSSYSHVWLKEVRFF 179
Db 142 FFWATSRALCSPPACSAFPRTGTSKQDVCKTLCTROPFSLAREACRSYSHVWLKEVRFF 201

Qy 180 NLQVLYPLLSDDPALNLRIVHLVRDPRVAVLSRQVATKALARDNGIVLGTNGKWEADPHL 239
Db 202 NLQVLYPLLSDDPALNLRIVHLVRDPRVAVLSRQVATKALARDNGIVLGTNGKWEADPHL 261

Qy 240 RVNVEVCRSHVRIAEALHKPPFPLODRYRLVRYEDLARDPLTVIRELYAFTGLTLPOL 299
Db 262 RLIREVCRSHVRIAEALHKPPFPLODRYRLVRYEDLARDPLTVIRELYAFTGLTLPOL 321

Qy 300 QTWHNTHGSGPGARREAFKTTSRDALSVSQAWRHLPFAKIRRVQELCGALQLGYR 359
Db 322 EAWHNTHTGSGIGKPIEAFTSSRNARNVSVQAWRHLPFAKIRRVQELCGALQLGYR 381

Qy 360 SVHSELEQRDLSDLLPRGMDSPKWS 387
Db 382 PVYSADQQRDLTLDVLRPGDPHFWSAS 409

RESULT 5
ID Q6GN39 PRELIMINARY; PRT; 392 AA.
AC Q6GN39;
DT 05-JUL-2004 (TRENBLrel. 27, Created)
DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)
DE MGC83048 protein.
GN Name=MGC83048;
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
ON NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=2238257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B.,
RA Brownstein M.J., Uedin T.B., Toshiyuki S., Carninci F., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,

RA Krzywinski M.I., Skalska U., Smalhus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16999-16903(2002).
RN [2]
RN SEQUENCE FROM N.A.
RP TISSUE=Embryo;
RC MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;
RX Klein S.L., Strausberg R.D., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.;
RA "Genetic and genomic tools for Xenopus research: The NIH Xenopus
RT initiative.";
RL Dev. Dyn. 225:384-391(2002).
RN [3]
RN SEQUENCE FROM N.A.
RP TISSUE=Embryo;
RC Klein S., Gerhard D.S.;
RA Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
RL EMBL: BC073681; AAH73681.1; -;
DR GO: GO:0008146; P:sulfotransferase activity; IEA.
DR InterPro: IPR000863; Sulfotransferase.
DR Pfam: PF00685; Sulfotransfer_1; 1.
SQ SEQUENCE 392 AA; 45969 MW; 7712810F8B4704D7 CRC64;

Query Match 57.7%; Score 1187; DB 2; Length 392;
Best Local Similarity 56.6%; Pred. No. 1.2e-93;
Matches 219; Conservative 77; Mismatches 92; Indels 4; Gaps 3;

QY 8 STVMSLLMVQGTGILVLSRQ--VPSPAGLGERVHVLSSWRSGSFVQGFQHPD 65
DB NVTVAGFLLQITFLIILYSRHTVLPDTEKT-EKVHLLLSWRSGSFVQGFQHPD 66
QY 66 VFVLMPEAHVWDTLSCGSPALHMAVRDLIRSVFLCDMDVFDAYLPWRRNISDLFQWAV 125
DB VFVLMPEAHVWMSFMFQNNKAVLHMAVRDLIRSVFLCDMSVFDAYIPNKNVSELFQWAV 126
QY 126 SRALCSPPVCEAFAGNISSEEVCKPLCATRPGLAQACSSYSHVVLKEVRFNQLVLY 105
DB SRALCSVPACSHDFREAITNCTVCKIKNPNFSKEESNTSHVILKEVRFNQLVLY 186
QY 186 PLASDPAALNRIHVLDRPRAVLRSEQTAKALARDNGIVLGTWVEADPRLRVNVEV 245
DB PLLTDPSSLNKKIHLVRDPAVAKRSQAMKYLTRDNGIVLNTNGTKID-DIRYDVMREI 245
QY 246 CRSHVRIAEALHKKPPFPLODRYLRVYEDLARDPLTVIRELYAFTGLTPOLOQWIHN 305
DB CQSHAQMYETAMDKAPSLKGRYMLVRYEDVDRDPLREINQMYEFANLKLTAKLKNWFYN 305
QY 306 ITHSGPGARREAFKTTSDALSVSQAWRHTLPFAKIRRVQELCGGALQLLYRSVHSEL 365
DB ITHGVGPGTKKEBFQTTSRNAVNSQAWRKDLSPQKIQTKIQTKISSEMNLLGYQFIDSEK 365
QY 366 EQRDLSDLLPLPGMDSKFKNWASSTEQ 392
DB ERKDMNDVLPKRNQYQFSLPNKEKK 392

RESULT 6
ID QBNCGS PRELIMINARY; PRT; 386 AA.
AC QBNCGS;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DE Hypothetical protein FLJ90265.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RN SEQUENCE FROM N.A.
RP Isogai T., Ota T., Nishikawa T., Hayashi K., Otsuki T., Sugiyama T.,

RA Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Kojima S., Nagahari K.,
RA Masuho Y., Ono T., Okano K., Yoshikawa Y., Aotsuka S., Sasaki N.,
RA Hattori A., Okumura K., Iwayanagi T., Ninomiya K.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK074746; BAC11177.1; -;
DR Genew; HGNC:1972; CHST4.
DR GO: GO:0008146; P:sulfotransferase activity; IEA.
DR GO: GO:0016740; F:transferase activity; IEA.
DR InterPro: IPR000863; Sulfotransferase.
DR Pfam: PF00685; Sulfotransfer_1; 1.
KW Transferase.
SQ SEQUENCE 386 AA; 45160 MW; 861869348319E42A CRC64;

Query Match 49.7%; Score 1022.5; DB 2; Length 386;
Best Local Similarity 56.7%; Pred. No. 1.8e-79;
Matches 216; Conservative 42; Mismatches 112; Indels 11; Gaps 7;

QY 1 MRLPRESSTVMSLLMVQGTGILV--LVSROVPS-SPAGLGERVHVLSSWRSGSFV 56
DB 1 MLLPK--KMKLLLFVSQMAILALPFHMYSHNLSLSKMAQPERMHVLSWRSGSFV 58
QY 57 GOLFSQHPDVFYLMPEAHVWDTLSCGSPALHMAVRDLIRSVFLCDMDVFDAYL-PWRR 115
DB GOLFGQHPDVFYLMPEAHVWMTFKQSTAWMLHMAVRDLIRAVFLCDMSVFDAYMEPGR 118
QY 116 NISDLFQWAVSRALCSPPVCEAFAGNISSEEVCKPLCATRPGLAQACSSYSHVVLKE 175
DB 119 ROSSLFQWNSRALCSAPACDIIPDEIIPRAHCRLLCSQPPFVVEKACRSYSHVVLKE 178
QY 176 VRFNQLVLYPLSDPALNRIHVLDRPRAVLRSEQTAKALARDNGIVLGTWTV-E 234
DB 179 VRFNQLQSLYPLKDPSSLNHLVHVRDPAVRSRERTKGDLMIDNRVIMGQHEQKLKK 238
QY 235 ADPLRVNNEVCRSVRIAEALHKKPPFPLODRYLRVYEDLARDPLTVIRELYAFTGLG 294
DB EQPYYVMQVICQSQLEIYK-TIQSLPKALQERYLLVRYEDLARADVAQTSRMYEFVGL 297
QY 295 LFPOLQWTHNTHSGPGARREAFKTTSDALSVSQAWRHTLPFAKIRRVQELCGGALQ 354
DB 298 FLUPHQTQWTHNTRGKMG--DHAFHTNARDALNSQAWRWSLPYEKVSRLQKACGDAMN 355
QY 355 LLGYRSVHSELEQDLSLDLL 375
DB 356 LLGYRHRVSRSEQRNLLDLL 376

RESULT 7
ID QBYSR3 PRELIMINARY; PRT; 386 AA.
AC QBYSR3;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE N-acetylglucosamine 6-O-sulfotransferase (L-selectin ligand
DE sulfotransferase GST-3).
GN Names=GST3;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RN SEQUENCE FROM N.A.
RP TISSUE=Tonsil;
RC MEDLINE=99264336; PubMed=10330415; DOI=10.1083/jcb.145.4.899;
RX Bistrup A., Bhakta S., Lee J.K., Belov Y.Y., Gunn M.D., Zuo F.R.,
Huang C.C., Kannagi R., Rosen S.D., Hemmerich S.;
RT "Sulfotransferases of two specificities function in the reconstitution
RT of high endothelial cell ligands for L-selectin.";
RL J. Cell Biol. 145:899-910(1999).
RN [2]
RN SEQUENCE FROM N.A.
RP TISSUE=Tonsil;

RX MEDLINE=21096027; PubMed=1181564; DOI=10.1093/glycob/11.1.1.75;
 RA Hemmerich S., Lee J.-K., Bhakta S., Bistrup A., Ruddle N.R.,
 RA Rosen S.D.;
 RA "Chromosomal localization and genomic organization of the galactose/N-
 RT acetylglucosamine/N-acetylglucosamine 6-O-sulfotransferase gene
 RT family";
 RL Glycobiology 11:75-87(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21332592; PubMed=11439191; DOI=10.1016/S0092-8674(01)00394-4;
 RA Yen J.-C., Hiraoka N., Petryniak B., Nakayama J., Ellices L.G.,
 RA Rabuka D., Hindsgaul O., March J.D., Lowe J.B., Fakuda M.;
 RT "Novel sulfated lymphocyte homing receptors and their control by a
 RT Core1 extension beta 1.3-N-acetylglucosaminyltransferase";
 RL Cell 105:957-969(2001).
 DR EMBL; AF131235; AA033015.1; -;
 DR EMBL; AF280088; AAG48246.1; -;
 DR EMBL; AF149783; AAK48417.1; -;
 DR GO; GO:0008146; P:cell adhesion; TAS.
 DR GO; GO:0007155; P:cell adhesion; TAS.
 DR GO; GO:0006928; P:cell motility; TAS.
 DR GO; GO:0007267; P:cell-cell signaling; TAS.
 DR GO; GO:0006955; P:immune response; TAS.
 DR GO; GO:0006477; P:protein amino acid sulfation; TAS.
 DR InterPro; IPR000863; Sulfotransferase.
 DR Pfam; PF00685; Sulfotransferase.
 KW Selectin; Transferrase.
 SQ SEQUENCE 386 AA; 45133 MW; 0C3BB4022417143A CRC64;

Query Match 49.5%; Score 1017.5; DB 2; Length 386;
 Best Local Similarity 56.4%; Pred. No. 4.7e-79;
 Matches 215; Conservative 43; Mismatches 112; Indels 11; Gaps 7;

Qy 1 MRLPRFSTVMSLLMVTQTILVF---LVSRQVPS-SPAGLGERVHVLLSSWRSGSFV 56
 Db 1 MLLPK--KMKLLFLVSMQAILALFFHMYSHNSSLKMAQPERMHHVLLSSWRSGSFV 58
 Qy 57 GOLFQHPDVPYLMPEAHVWDTLSQGSAPALHMAVRDLRSVFLCDMDVFDAYL-PWRR 115
 Db 59 GOLFQHPDVPYLMPEAHVWMTFKQSTAWMLHMAVRDLRAVFLCDMSVFDAYMEGPR 118
 Qy 116 NISDLFQAVSRALCSPVPCAFARGNISSEVCKPLCATRPFGLAQACSSYSHVVLKE 175
 Db 119 QSSSLFQWNSRALCSAPACDIIPQDEIIPRAHCRLLCSQPFVEVKEACRSYSHVVLKE 178
 Qy 176 VRFNQLVPLSDPALNLRIHVLPDRAVLSRQETAKALARDNGIVLGTNGTWV-E 234
 Db 179 VRFNQLVPLSDPALNLRIHVLPDRAVLSRQETAKALARDNGIVLGTNGTWV-E 238
 Qy 235 ADPLRVNVECRSHVRIAEALHKKPPFPDQRYLVRYEDLARDPLTVIRELYAFTGLG 294
 Db 239 EDQPVYVQVTCOSQLEIYK-TIQLSPKALQERLVRYEDLARAPVAQTSRMTEFVGL 297
 Qy 295 LTPOLQTHNITHGSGPGARREAFKTSRDLVSQARHTLPFAKIRRVQELCGGALQ 354
 Db 298 FLPHLQTVHNTIRGKMG--DHAFHTNARDALNVQAWRSLPVEKVSRLQKACGDAMN 355
 Qy 355 LLGYRSHVSELEQRLDLSL 375
 Db 356 LLGYRSHVSELEQRLDLSL 376

RESULT 8
 Q81V46 PRELIMINARY; PRT; 370 AA.
 AC Q81V46;
 DT 01-MAR-2003 (TRENBLrel. 23, Created)
 DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
 DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
 DE CHST4 protein.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J.M., Haileh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Sapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Pahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RA Strausberg R.;
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC035282; AAH35282.1; -;
 DR GO; GO:0008146; P:sulfotransferase activity; IEA.
 DR InterPro; IPR000863; Sulfotransferase.
 DR Pfam; PF00685; Sulfotransferase.
 SQ SEQUENCE 370 AA; 43249 MW; EA8D76EA4E73C625 CRC64;

Query Match 49.1%; Score 1010.5; DB 2; Length 370;
 Best Local Similarity 60.2%; Pred. No. 1.8e-78;
 Matches 204; Conservative 38; Mismatches 92; Indels 5; Gaps 4;

Qy 39 ERHVLLVLSWRSGSFVQGLFSQHDPVFLYLMPEAHVWDTLSQGSAPALHMAVRDLRS 98
 Db 25 ERHVLLVLSWRSGSFVQGLFSQHDPVFLYLMPEAHVWMTFKQSTAWMLHMAVRDLRS 84
 Qy 99 VFICDMVFDAYL-PWRRNISDLFQAVSRALCSPVPCAFARGNISSEVCKPLCATR 157
 Db 85 VFICDMVFDAYMEPGPRROSSLFPQWNSRALCSAPACDIIPQDEIIPRAHCRLLCSQ 144
 Qy 158 FGLAQACSSYSHVVLKEVRFNQLVPLSDPALNLRIHVLPDRAVLSRQETAKA 217
 Db 145 FEVVEKACRSYSHVVLKEVRFNQLVPLSDPALNLRIHVLPDRAVLSRQETAKA 204
 Qy 218 LARDNGIVLGTNGTWV-EADPLRVNVECRSHVRIAEALHKKPPFPDQRYLVRYEDL 276
 Db 205 LMDSIRVNGQHEQKLKEDQPVYVQVTCOSQLEIYK-TIQLSPKALQERLVRYEDL 263
 Qy 277 ARPLTVIRELYAFTGLTLPOLQTHNITHGSGPGARREAFKTSRDLVSQARHT 336
 Db 264 ARAPVAQTSRMTEFVGLFPLHQLTVHNTIRGKMG--DHAFHTNARDALNVQAWR 321
 Qy 337 LPFAKIRRVQELCGGALQQLGYRSHVSELEQRLDLSL 375
 Db 322 LPYKVSRLQKACGDAMNLLGYRSHVSELEQRLDLSL 360

RESULT 9
 Q9R111 PRELIMINARY; PRT; 388 AA.
 AC Q9R111;
 DT 01-MAY-2000 (TRENBLrel. 13, Created)
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
 DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)
 DE L-selectin ligand sulfotransferase (Chst4 protein).

| | |
|---|--|
| GN | Name=Chst4; |
| OS | Mus musculus (Mouse). |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; |
| OX | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. |
| NCBI_TaxID=10090; | |
| RN | [1] |
| RN | SEQUENCE FROM N.A. |
| RP | MEDLINE=99361934; PubMed=10435581; DOI=10.1016/S1074-7613(00)80083-7; |
| RX | Hiraka N., Petryniak B., Nakayama J., Teuboi S., Suzuki M., Yeh J.C., |
| RA | Izawa D., Tanaka T., Miyasaka M., Lowe J.B., Fukuda M.; |
| RT | "A novel, high endothelial venule-specific sulfotransferase expresses |
| RL | 6-sulfo sialyl Lewis(x), an L-selectin ligand displayed by CD34."; |
| RL | Immunity 11:79-89(1999). |
| RP | [2] |
| RN | SEQUENCE FROM N.A. |
| RP | STRAIN=FVB/N; TISSUE=Colon; |
| RC | MEDLINE=22388857; PubMed=12477932; DOI=10.1073/pnas.242603899; |
| RX | Straussberg R.L., Feingold E.A., Grouse L.H., Berge J.G., |
| RA | Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., |
| RA | Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., |
| RA | Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J.J., Heieh F., |
| RA | Diatchenko L., Marushina K., Farmer A.A., Rubin G.M., Hong L., |
| RA | Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., |
| RA | Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C., |
| RA | Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J., |
| RA | Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., |
| RA | Richards S., Worthy K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., |
| RA | Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., |
| RA | Fahey J., Helton E., Ketteman M., Madan A., Rodriguez S., Sanchez A., |
| RA | Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., |
| RA | Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C., |
| RA | Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S., |
| RA | Krzywinski M.I., Skalska U., Smalish D.E., Schnertch A., Schein J.E., |
| RA | Jones S.J., Marra M.A.; |
| RA | "Generation and initial analysis of more than 15,000 full-length human |
| RT | and mouse cDNA sequences." |
| RL | Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002). |
| RP | [3] |
| RN | SEQUENCE FROM N.A. |
| RP | STRAIN=FVB/N; TISSUE=Colon; |
| RC | Straussberg R.; |
| RL | Submitted (SEP-2003) to the ENBL/GenBank/DBJ databases. |
| DR | EMBL; AF109155; RAD45579.1; - |
| DR | EMBL; BC057886; AAH57886.1; - |
| DR | MGD; MG1:1349479; Chet4. |
| DR | GQ; GO:0016021; C:integral to membrane; TAS. |
| DR | InterPro; IPR000863; Sulfoltransferase. |
| DR | Pfam; PF00685; Sulfoltransfer_1. |
| KW | Selectin; Transferase. |
| SQ | SEQUENCE 388 AA; 44635 MW; 6D53771AFB6884AE CRC64; |
| | |
| Query Match | 48.2%; Score 991; DB 2; Length 388; |
| Best Local Similarity | 55.2%; Pred.No. 9.1e-77; |
| Matches 208; Conservative 43; Mismatches 114; Indels 12; Gaps | |
| | |
| QY | 2 RLPRF--SSTVMLSLLMVQTGILVLVSQRQVPSSPAGLGERVHVLTLSWSFGSGSVGQLF |
| Db | 8 RLFLMGQSIVTVALFIHMSVHRHLQSRESRP-----VHVLTLSWSFGSGSVGQLF |
| QY | 61 SQHPDPVYLMEPAFWHDPTLSQGSAPALHMAVRDLIRSFVLCDMDVDAYL-PWRNRISD |
| Db | 62 GQHPDPVYLMEPAFWHMVTFTSTAWKLHMAVRDLLRSVFLCDMSVFDAYMNPGRKQSS |
| QY | 120 LFQAVNSRALCSPPVCEAPARGNISSEBEVKPLCATRPFGLQAQCACSSYSHVVLKEVRF |
| Db | 122 LFQWEQSRALCAPVCDFFPAAHEIISPCHKLLCCQQPFDMWVEKACRSHGFVVLKEVRF |
| QY | 180 NLQVLYPLLSDPALNLRIVHLVRDPAVLRSREQTAKALARDNIGIVLGTN-GTWVADPR |
| Db | 182 SIQALYPLLTDSPLNLVHLVRDPAVFRGREHTTELIVVDSHIVLQGHLETIKEDQP |
| QY | 239 LRVMNEVCRSHVRIAEEALHKPPPPLQDRYLRVLEDLARLDPLTVIRELYAFTGLGLTPQ |

[illegible]

RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.,
 RT "RIKEN integrated sequence analysis (RISA) system-384-format
 RT sequencing pipeline with 384 multicapillary sequencer.";
 RL Genome Res. 10:1757-1771(2000).
 RN [7]
 RN SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Tongue;
 RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
 RA Arakawa T., Bono H., Carrinci P., Fukuda S., Fukunishi Y., Furuno M.,
 RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
 RA Inotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
 RA Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
 RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
 RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
 RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
 RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
 RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
 RA Muramatsu M., Hayashizaki Y.,
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF131236; AAB33016.1; -
 DR EMBL; AK009113; BAB26078.1; -
 DR MGD; MGI:1349479; Chat4.
 DR GO; GO:0016021; C:integral to membrane; TAS.
 DR InterPro; IPR000863; Sulfotransferase.
 DR Pfam; PF00685; Sulfotransfer_1; 1.
 KW Transferase.
 SQ SEQUENCE 388 AA; 44694 MW; D1E9D7796DF8574D CRC64;

Query Match 47.8%; Score 984; DB 2; Length 388;
 Best Local Similarity 54.9%; Pred. No. 3.7e-76;
 Matches 207; Conservative 44; Mismatches 114; Indels 12; Gaps 6;

QY 2 RLPRF-SSTVMSLLVMQTGILVLSRQVPSPPAGLGERVHVLVLSWSRSGSFVGOLF 60
 DB 8 RLLMFLGQVIVVAFIHMSVHRHLSQREESRRP-----VHVLVLSWSRSGSFVGOLF 61
 QY 61 SQHPDVFYLMPEPAWVMDTISQGSAPALHMAVRDLIRSVFLCDMDVFDAYL-PWRRNISD 119
 DB 62 GQHPDVFYLMPEPAWVMDTISQGSAPALHMAVRDLIRSVFLCDMDVFDAYL-PWRRNISD 121
 QY 120 LFQAVSRALCPVCEAFAGNITSSEVCKPLCATRPFGLAQACSSYSHVVLKEVRF 179
 DB 122 LFQEQSRALCPVCEAFAGNITSSEVCKPLCATRPFGLAQACSSYSHVVLKEVRF 181
 QY 180 NLQVLPDLPALNLRVHVRDPAVLRSEQTAKALARDNGIVLGTN-GTWVEADPR 238
 DB 182 SLQALYPLLTDPNLNHHVHVRDPAVLRSEQTAKALARDNGIVLGTN-GTWVEADPR 241
 QY 239 LRVNVECRSHVRIAEALHKKPPFLQDRYRLRYEDLARDPLTVIRELYAFTGLTPQ 298
 DB 242 YYAMKICKSQDVIVK-AIQLPEALQORYFLRYEDLARDPLTVIRELYAFTGLTPQ 300
 QY 299 LOTWIHNTHSGGPGARREAFKTSRDALSQAWRHRLTPAKIRRVQELCGGALLGY 358
 DB 301 LOTWYVYVTRGKMG--QHAFTNARNALNVSQAWRWSLPYKVSQLODACEAMDLLGY 358
 QY 359 RSVHSELEQDLSLDLL 375
 DB 359 LQVRSQEQGNLSLDLL 375

RESULT 11
 Q794G9 PRELIMINARY; PRT; 483 AA.
 ID Q794G9;
 AC Q794G9;
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
 DE N-acetylglucosamine-6-O-sulfotransferase (N-Acetylglucosamine 6-O-

DE Sulfotransferase-1).
 OS Name=Gn6st-1;
 GN Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC TISSUE=Whole embryos;
 RX MEDLINE=98380482; PubMed=9712885; DOI=10.1074/jbc.273.35.22577;
 RA Uchimura K., Muramatsu H., Kadomatsu K., Fan Q., Kurosawa N.,
 RA Mitsuoka C., Kannagi R., Habuchi O., Muramatsu T.,
 RT "Molecular cloning and characterization of an N-acetylglucosamine-6-O-
 RT sulfotransferase.";
 RL J. Biol. Chem. 273:22577-22583(1998).
 RN [2]
 RN SEQUENCE FROM N.A.
 RA Uchimura K., Kadomatsu K., El-Pasakhany F.M., Singer M.S., Izawa M.,
 RA Kannagi R., Takeda N., Rosen S.D., Muramatsu T.,
 RT "N-Acetylglucosamine 6-O-Sulfotransferase-1 Regulates Expression of L-
 RT Selectin Ligands and Lymphocyte Homing.";
 RL J. Biol. Chem. 279:35001-35008(2004).
 DR EMBL; AB011452; BAA32139.1; -
 DR EMBL; AB125058; BAA32137.1; -
 DR GO; GO:0005615; C:extracellular space; TAS.
 DR GO; GO:0008146; F:sulfotransferase activity; IDA.
 DR GO; GO:0006044; P:N-acetylglucosamine metabolism; TAS.
 DR InterPro; IPR000863; Sulfotransferase.
 DR Pfam; PF00685; Sulfotransfer_1; 1.
 KW Transferase.
 SQ SEQUENCE 483 AA; 52830 MW; 831FA08F5FEFD70E CRC64;

Query Match 31.1%; Score 640; DB 2; Length 483;
 Best Local Similarity 36.2%; Pred. No. 2.1e-46;
 Matches 145; Conservative 73; Mismatches 118; Indels 64; Gaps 10;

QY 30 VPSSPAG-----LG-----ERVHVLVLSWSRSGSFVGOLFQHPDVFYLM 71
 DB 89 VPAAGSAGAAASLGNATRGTRGGGDKQLVYVFTTWSRSGSFPGELFNQNPVFFLYE 148
 QY 72 PAHVMDTISQGSAPALHMAVRDLIRSVFLCDMDVFDAYL-P---WRRNISD---FQWAVS 126
 DB 149 PVHVMWQKLYPGDAVSLQGAARDMLSAFYRCDLSVFLYSPAGSGGRNLTTLGTFGAATN 208
 QY 127 RALCSPVCEAFAG--NISSEVCKPLCATRPFGLAQACSSYSHVVLKEVRFNQLV 184
 DB 209 KVCSSPLCPAYRKEVVGVLVDVCKK-CPQRLARFEECKRYTVVKGVRVFDVAVL 267
 QY 185 YPLSDPDLPALNLRVHVRDPAVLRSEQTAKALARDNGIVLGTN-GTWVEADPR----- 238
 DB 268 APLLKDPALDLPALNLRVHVRDPAVLRSEQTAKALARDNGIVLGTN-GTWVEADPR 321
 QY 239 -----LRVNEVCRSHVRIAEALHKKPPFLQDRYRLRYEDLARDPLTVIRELYAFT 275
 DB 322 LEAAGHKLGAKEGGMGPADYHALGAMVEVCNSAKTLQAL-QPPDMLQGHLYLVRYED 380
 QY 276 LARDPLTVIRELYAFTGLTLPQLOTWIHNTHSGGPGARREAFKTSRDALSQAWRH 335
 DB 381 LVGDPVKTLLRVYDFVGLLVSPMEQFALNMTSGSGSSK--PFWARNATQANAWRT 438
 QY 336 TLPEAKIRRVQELCGGALLQLYGRSVHSELEQDLSLDLL 375
 DB 439 ALTFQQLKQVEEFCYQPMVLYGVVNSPEEVKDSKTL 478

RESULT 12
 O88276 PRELIMINARY; PRT; 530 AA.
 ID O88276;
 AC O88276;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)

DE N-acetylglucosamine-6-O-sulfotransferase long form.
GN Names=Chst2;
OS Mus musculus (Mouse).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Whole embryos;
RX MEDLINE=9830482; PubMed=9712885; DOI=10.1074/jbc.273.35.22577;
RA Uchimura K., Muramatsu H., Kadomatsu K., Fan Q., Kurosawa N.,
RA Mitsuoka C., Kannagi R., Habuchi O., Muramatsu T.;
RT "Molecular cloning and characterization of an N-acetylglucosamine-6-O-
RT sulfotransferase";
RL J. Biol. Chem. 273:22577-22583 (1998).
DR ENBL; AB011452; BAA32138.1; -;
DR MGD; MGI:1891160; Chst2.
DR GO; GO:0005615; C:extracellular space; TAS.
DR GO; GO:0008146; P:sulfotransferase activity; IDA.
DR GO; GO:0006044; P:N-acetylglucosamine metabolism; TAS.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
KW Transferase.
SQ SEQUENCE 530 AA; 57814 MW; A113E1B735C363EC CRC64;

Query Match 31.1%; Score 640; DB 2; Length 530;
Best Local Similarity 36.2%; Pred. No. 2.4e-46;
Matches 145; Conservative 73; Mismatches 118; Indels 64; Gaps 10;

QY 30 VPSSPAG-----LG-----ERHVLVLSWSRSGSFVGLFSQHPDVFYLM 71
DB 136 VPAAGSAGAAASLGNATRGGGDKRLVYVFTTWSRSGSFGLFNQNEVFVLYE 195
QY 72 PAHVVDTLTSGSAPALHMAVRDLIRSVFLCDMDVFDAYLP---WRNISDL--FQWAVS 126
DB 196 PVHVMQKLYPGDAVSLQGAARDMLSAFYRCDLSVFLYSPAGSGGRNLTTLGFGAATN 255
QY 127 RALCSPPVCEAFARG--NISSEVCKPLCATRPFGLAQEACSSYSHVLKEVFFNLQVL 184
DB 256 KVCSSPLCPAYRKEVVGVLVDVCKK-CPQRLARFEBCRYKRVVIGKRVFVAVL 314
QY 185 YLLSDPALNLRIHVLDRPRAVLRSREQTAKALARDNGIVLGTNGTWVEADPR----- 238
DB 315 APLKDPALDLKVIHLVRDPRAVASSRIRSRHGLIRESLQVVRSR-----DPAHRMPF 368
QY 239 -----LRVNEVCRSHVRIAEALHKPPFPLODRVLYRYED 275
DB 369 LEAAGHKLGAKEGMGPADYHALGAMEVICNSMAKTLOTAL-QPPDWLQGHVLYRYED 427
QY 276 LARDPLTVIRELYAFTGLTPOLOTWIHNIHSGSGPGARREAFKTTSRDALSVQAWRH 335
DB 428 LVGDPVKTLRRVDFVGLLVSPMEQFALNMWTSGSGSSK--PFVVSARNATQAANWRT 485
QY 336 TLPFAKIRVQELCGGALQLLYGRSVHSELEORDLSLDLL 375
DB 486 ALTFQIQKVEEFCYQPMVGLGVYRVSPEEVKDLSTLL 525

RESULT 13
Q80WV3 PRELIMINARY; PRT; 530 AA.
AC Q80WV3;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Chst2 protein.
GN Names=Chst2;
OS Mus musculus (Mouse).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.F., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toehiyuki S., Carninci P., Prange C.,
RA Rohn S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Morzy D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Medan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skaleka U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strausberg R.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR ENBL; BC051963; AAH51963.1; -;
DR MGD; MGI:1891160; Chst2.
DR GO; GO:0005615; C:extracellular space; TAS.
DR GO; GO:0008146; P:sulfotransferase activity; IDA.
DR GO; GO:0006044; P:N-acetylglucosamine metabolism; TAS.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
SQ SEQUENCE 530 AA; 57828 MW; 275363BF15440730 CRC64;

Query Match 31.0%; Score 637; DB 2; Length 530;
Best Local Similarity 36.0%; Pred. No. 4.3e-46;
Matches 144; Conservative 74; Mismatches 118; Indels 64; Gaps 10;

QY 30 VPSSPAG-----LG-----ERHVLVLSWSRSGSFVGLFSQHPDVFYLM 71
DB 136 VPAAGSAGAAASLGNATRGGGDKRLVYVFTTWSRSGSFGLFNQNEVFVLYE 195
QY 72 PAHVVDTLTSGSAPALHMAVRDLIRSVFLCDMDVFDAYLP---WRNISDL--FQWAVS 126
DB 196 PVHVMQKLYPGDAVSLQGAARDMLSAFYRCDLSVFLYSPAGSGGRNLTTLGFGAATN 255
QY 127 RALCSPPVCEAFARG--NISSEVCKPLCATRPFGLAQEACSSYSHVLKEVFFNLQVL 184
DB 256 KVCSSPLCPAYRKEVVGVLVDVCKK-CPQRLARFEBCRYKRVVIGKRVFVAVL 314
QY 185 YLLSDPALNLRIHVLDRPRAVLRSREQTAKALARDNGIVLGTNGTWVEADPR----- 238
DB 315 APLKDPALDLKVIHLVRDPRAVASSRIRSRHGLIRESLQVVRSR-----DPAHRMPF 368
QY 239 -----LRVNEVCRSHVRIAEALHKPPFPLODRVLYRYED 275
DB 369 LEAAGHKLGAKEGMGPADYHALGAMEVICNSMAKTLOTAL-QPPDWLQGHVLYRYED 427
QY 276 LARDPLTVIRELYAFTGLTPOLOTWIHNIHSGSGPGARREAFKTTSRDALSVQAWRH 335
DB 428 LVGDPVKTLRRVDFVGLLVSPMEQFALNMWTSGSGSSK--PFVVSARNATQAANWRT 485
QY 336 TLPFAKIRVQELCGGALQLLYGRSVHSELEORDLSLDLL 375
DB 486 ALTFQIQKVEEFCYQPMVGLGVYRVSPEEVKDLSTLL 525

RESULT 14
Q9UED5
ID Q9UED5
AC Q9UED5;
PRELIMINARY; PRT; 483 AA.

01-MAY-2000 (TrEMBLrel. 13, Created)
01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
N-acetylglucosamine-6-O-sulfotransferase (GlcNAc6ST).
Name=GN6ST;
Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
RP TISSUE=Brain;
RC MEDLINE=98391845; PubMed=9722682;
RA Uchimura K., Muramatsu H., Kaname T., Ogawa H., Yamakawa T., Fan Q.,
RA Mitsuoka C., Kannagi R., Habuchi O., Yokoyama I., Yamamura K.,
RA Ozaki T., Nakagawara A., Kadamatsu K., Muramatsu T.;
"Human N-acetylglucosamine-6-O-sulfotransferase involved in the
RT biosynthesis of 6-sulfo sialyl Lewis X: molecular cloning, chromosomal
RT mapping, and expression in various organs and tumor cells.";
RL J. Biochem. 124:670-678 (1998).
[2]
SEQUENCE FROM N.A.
RP TISSUE=Placenta;
RC MEDLINE=20498786; PubMed=11042394; DOI=10.1016/S0304-4165(00)00136-7;
RA Sakaguchi H., Kitagawa H., Sugahara K.;
RT "Functional expression and genomic structure of human N-
RT acetylglucosamine-6-O-sulfotransferase that transfers sulfate to b-N-
RT acetylglucosamine at the nonreducing end of an N-acetylglucosamine
RT sequence.";
RL Biochim. Acta 1523:269-276 (2000).
DR EMBL; AB014679; BAA34265.2; -
DR EMBL; AB021124; BAB16886.1; -
DR EMBL; AB021125; BAB16887.1; -
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
KW Transferase.
SQ SEQUENCE 483 AA; 52787 MW; E27797D44931BA18 CRC64;
Query Match 30.9%; Score 636.5; DB 2; Length 483;
Best Local Similarity 37.1%; Pred. No. 4.2e-46;
Matches 140; Conservative 73; Mismatches 117; Indels 47; Gaps 9;
Qy 36 GLGE-RVHVLVLSWRSGSFVQGLFSQHPDVFYLMPEAWHVDLTSQGSAPALHMAVRD 94
Db 112 GVGDKRQLVYVFTTWRSGSFFGELFNQNPVEFVYFVHVMQKLYPGDAVSLQGAARD 171
Qy 95 LIRSVFLCDMDVFDAYLP---WRRNISDL--FQWAVSRALCSPVCEAFARG--NISSEE 147
Db 172 MLSALYKCDLSVFLYSPAGSGGRNLTTLGIFGAATNKVVCSSPLCPAYRKEVVGVDPR 231
Qy 148 VCKPLCATRPFGLAQEACSSYSHVLKVEFFNQLVPLLSDPALNLRIVHLVRDPRV 207
Db 232 VCKK-CPQRLARFECECRKRYTLVIGRVFVAVLAPLRLDPAALDKVHLVRDPRV 290
Qy 208 LRSREQTAKALRNGIVLGTNGTWVEADPR----- 238
Db 291 ASSRIRSRHGLIRESLQVRSR-----DPAHRMPFLEAGHKLGAKEGVGGPADYHA 344
Qy 239 LRVNVECRSHVRIAEALHKPPFPLODRYRLVRYEDLARDPLTVIRELYAFTGLTTPQ 298
Db 345 LGAMEVICNSMAKTLQAL-OPDNLQGHVLYVRYEDLVGDPVKTLLRRVDFVGLLVSP 403
Qy 299 LQTVIHNTHGSGPGARREAFKTTSDALSVSQAWRHTLPFAKIRRYQELCGGALLQ 358
Db 404 MEQFALNMTSGSGSSK--PFVVSARNATQAANARWTALTQQIKQVEEFCYQPMVGLY 461
Qy 359 RSVHSELEQRDLSDLL 375
Db 462 ERVNSPEVKDLSKTL 478

RESULT 15
QYV4C5 PRELIMINARY; PRT; 530 AA.
ID QYV4C5; Q9GZNS; Q9V6P2;
AC QYV4C5; Q9GZNS; Q9V6P2; 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 27, Last annotation update)
DE Long form of N-acetylglucosamine-6-O-sulfotransferase (GlcNAc6ST)
DE (Carbohydrate sulfotransferase 2).
DE Name=GN6ST; Synonyms=CHST2;
GN Name=GN6ST; Synonyms=CHST2;
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
RP TISSUE=Brain;
RC MEDLINE=98391845; PubMed=9722682;
RA Uchimura K., Muramatsu H., Kaname T., Ogawa H., Yamakawa T., Fan Q.,
RA Mitsuoka C., Kannagi R., Habuchi O., Yokoyama I., Yamamura K.,
RA Ozaki T., Nakagawara A., Kadamatsu K., Muramatsu T.;
"Human N-acetylglucosamine-6-O-sulfotransferase involved in the
RT biosynthesis of 6-sulfo sialyl Lewis X: molecular cloning, chromosomal
RT mapping, and expression in various organs and tumor cells.";
RL J. Biochem. 124:670-678 (1998).
[2]
SEQUENCE FROM N.A.
RP TISSUE=Umbilical vein endothelium;
RC MEDLINE=99168906; PubMed=10049591; DOI=10.1006/geno.1998.5653;
RA Li X., Tedder T.F.;
RT "CHST1 and CHST2 sulfotransferases expressed by human vascular
RT endothelial cells: cDNA cloning, expression, and chromosomal
RT localization.";
RL Genomics 55:345-347 (1999).
DR EMBL; AB014680; BAA34266.2; -
DR EMBL; AF083066; AAD20981.1; -
DR Genew; HGNC:1970; CHST2.
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
KW Transferase.
SQ SEQUENCE 530 AA; 57857 MW; A82CA227B9D5651B CRC64;
Query Match 30.9%; Score 636.5; DB 2; Length 530;
Best Local Similarity 37.1%; Pred. No. 4.7e-46;
Matches 140; Conservative 73; Mismatches 117; Indels 47; Gaps 9;
Qy 36 GLGE-RVHVLVLSWRSGSFVQGLFSQHPDVFYLMPEAWHVDLTSQGSAPALHMAVRD 94
Db 159 GVGDKRQLVYVFTTWRSGSFFGELFNQNPVEFVYFVHVMQKLYPGDAVSLQGAARD 218
Qy 95 LIRSVFLCDMDVFDAYLP---WRRNISDL--FQWAVSRALCSPVCEAFARG--NISSEE 147
Db 219 MLSALYKCDLSVFLYSPAGSGGRNLTTLGIFGAATNKVVCSSPLCPAYRKEVVGVDPR 278
Qy 148 VCKPLCATRPFGLAQEACSSYSHVLKVEFFNQLVPLLSDPALNLRIVHLVRDPRV 207
Db 279 VCKK-CPQRLARFECECRKRYTLVIGRVFVAVLAPLRLDPAALDKVHLVRDPRV 337
Qy 208 LRSREQTAKALRNGIVLGTNGTWVEADPR----- 238
Db 338 ASSRIRSRHGLIRESLQVRSR-----DPAHRMPFLEAGHKLGAKEGVGGPADYHA 391
Qy 239 LRVNVECRSHVRIAEALHKPPFPLODRYRLVRYEDLARDPLTVIRELYAFTGLTTPQ 298
Db 392 LGAMEVICNSMAKTLQAL-OPDNLQGHVLYVRYEDLVGDPVKTLLRRVDFVGLLVSP 450
Qy 299 LQTVIHNTHGSGPGARREAFKTTSDALSVSQAWRHTLPFAKIRRYQELCGGALLQ 358
Db 451 MEQFALNMTSGSGSSK--PFVVSARNATQAANARWTALTQQIKQVEEFCYQPMVGLY 508
Qy 359 RSVHSELEQRDLSDLL 375

Db 509 ERVNSPEEVKDLKTL 525

Search completed: June 23, 2005, 08:48:57
Job time : 46.3472 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2005, 08:32:43 ; Search time 42.6885 Seconds
(without alignments)
3578.727 Million cell updates/sec

Title: US-10-697-828-7

Perfect score: 2057

Sequence: 1 MRLPFSSTVMSLLMVQTG.....LPRGMSFKWASSTKQPES 395

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

A_Geneseq_16Dec04:*

1: Geneseqp1980s:*

2: Geneseqp1990s:*

3: Geneseqp2000s:*

4: Geneseqp2001s:*

5: Geneseqp2002s:*

6: Geneseqp2003as:*

7: Geneseqp2003bs:*

8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|--------------------|
| 1 | 2057 | 100.0 | 395 | 4 | AAY72638 Mouse gly |
| 2 | 2057 | 100.0 | 395 | 5 | AAU11275 Murine in |
| 3 | 2042.5 | 99.3 | 418 | 5 | ABB81557 Mouse int |
| 4 | 1727 | 84.0 | 395 | 5 | ABB81555 Consensus |
| 5 | 1708 | 83.0 | 395 | 4 | AAY72640 Human gly |
| 6 | 1708 | 83.0 | 395 | 5 | ABB81554 Human cor |
| 7 | 1708 | 83.0 | 395 | 5 | AAE15438 Human dru |
| 8 | 1708 | 83.0 | 395 | 7 | ADI21086 Novel hum |
| 9 | 1708 | 83.0 | 395 | 8 | ADL61235 Human tyr |
| 10 | 1536.5 | 74.7 | 390 | 4 | AAY72639 Human gly |
| 11 | 1536.5 | 74.7 | 390 | 5 | ABB81556 Human int |
| 12 | 1436 | 69.8 | 418 | 3 | AAU11274 Human L-s |
| 13 | 1333 | 64.8 | 394 | 7 | ADJ70405 Human ORF |
| 14 | 1022.5 | 49.7 | 386 | 4 | AAU93309 Human pol |
| 15 | 1022.5 | 49.7 | 386 | 8 | ADL30784 Human pro |
| 16 | 1017.5 | 49.5 | 386 | 2 | AAU39918 Human gly |
| 17 | 1015.5 | 49.4 | 380 | 5 | AAU11274 Human L-s |
| 18 | 986.5 | 48.0 | 386 | 3 | AAY79219 Human tra |
| 19 | 984 | 47.8 | 388 | 2 | AAU39919 Mouse gly |
| 20 | 712 | 34.6 | 169 | 5 | ABB81558 Human cor |
| 21 | 677 | 32.9 | 169 | 5 | ABB81559 Human int |
| 22 | 647.5 | 31.5 | 484 | 2 | AAU31657 Human N-a |
| 23 | 647.5 | 31.5 | 531 | 5 | AAU69414 Lung snal |
| 24 | 647.5 | 31.5 | 531 | 8 | ADR14283 Human NF- |
| 25 | 640 | 31.1 | 483 | 2 | AAY31656 Mouse N-a |

| | | | | | | |
|----|-------|------|-----|---|----------|--------------------|
| 26 | 636.5 | 30.9 | 530 | 4 | AAB95367 | Aab95367 Human pro |
| 27 | 636.5 | 30.9 | 530 | 8 | ADQ18590 | Adq18590 Human sof |
| 28 | 579 | 28.1 | 486 | 6 | ABP56121 | Abp56121 Human cho |
| 29 | 563.5 | 27.4 | 479 | 7 | ABM85237 | Abm85237 Human pro |
| 30 | 563.5 | 27.4 | 499 | 6 | ABR41139 | Abr41139 Human DIT |
| 31 | 561.5 | 27.3 | 479 | 2 | AAW52863 | Aaw52863 Glycoami |
| 32 | 550 | 26.7 | 481 | 7 | ABM85236 | Abm85236 Mouse pro |
| 33 | 549 | 26.7 | 410 | 7 | ADN95522 | Adn95522 Human BEC |
| 34 | 549 | 26.7 | 411 | 2 | AAW61100 | Aaw61100 Keratan s |
| 35 | 549 | 26.7 | 411 | 5 | AAE25356 | Aae25356 Human cho |
| 36 | 549 | 26.7 | 411 | 6 | ABU03503 | Abu03503 Angiogene |
| 37 | 549 | 26.7 | 411 | 7 | ADJ68589 | Adj68589 Human hea |
| 38 | 549 | 26.7 | 411 | 8 | ADQ18725 | Adq18725 Human sof |
| 39 | 529.5 | 25.7 | 171 | 5 | ABB81560 | Abb81560 Human hig |
| 40 | 519 | 25.2 | 128 | 7 | ADI21571 | Adi21571 Novel hum |
| 41 | 513.5 | 25.0 | 504 | 8 | ABM85022 | Abm85022 Human dia |
| 42 | 495 | 24.1 | 458 | 2 | AAW06480 | Aaw06480 Chick cho |
| 43 | 377 | 18.3 | 169 | 5 | ABB81561 | Abb81561 Human N-a |
| 44 | 351.5 | 17.1 | 174 | 5 | ABB81563 | Abb81563 Human cho |
| 45 | 338 | 16.4 | 179 | 5 | ABB81562 | Abb81562 Human ker |

ALIGNMENTS

RESULT 1

AAU72638

ID AAY72638 standard; protein; 395 AA.

XX

AC AAY72638;

XX

DT 02-MAY-2001 (first entry)

XX

DE Mouse glycosyl sulfotransferase-4 (GST-4).

XX

KW Mouse; glycosyl sulfotransferase-4; GST-4; immunosuppressive; therapy;
KW selectin binding inhibitor; gene therapy; inflammation;
KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
KW polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenalitis;
KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
KW asthma; hypersensitivity; rheumatic fever; tissue rejection;
KW chromosome 8E1.

XX Mus musculus.

OS

XX WO200106015-A1.

XX

PD 25-JAN-2001.

XX

PF 19-JUL-2000; 2000WO-US019741.

XX

PR 20-JUN-1999; 99US-0144694P.

XX

PR 13-JUN-2000; 2000US-00593828.

XX

PA (REGC) UNIV CALIFORNIA.

XX

PI Rosen SD, Lee JK, Hemmerich S;

XX

DR WPI: 2001-138471/14.

XX

DR N-PSDB; AAD02696.

XX

PT New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for

diagnostic and therapeutic agent screening applications.

XX

PS Claim 3; Fig 2; 128pp; English.

XX

CC The present sequence is mouse glycosyl sulfotransferase-4 (GST-4). GST-4
CC gene is found on chromosome 8E1. GST is a type 2 membrane protein useful
CC for inhibiting a binding event between a selectin and a selectin ligand,
CC which comprises contacting the selectin with a non-sulphated selectin

CC ligand, GST and a small molecular agent that inhibits the sulphation
 CC activity of GST. GST is also useful in inhibiting a selectin mediated
 CC binding event. GST is useful in gene therapy to treat disorders such as
 CC acute or chronic inflammation, systemic lupus erythematosus (SLE),
 CC rheumatoid arthritis, polyarteritis nodosa, polymyositis,
 CC dermatomyositis, systemic sclerosis, diabetes, glomerulonephritis,
 CC myasthenia gravis, Sjogren's syndrome, Hashimoto's disease, Grave's
 CC disease, adrenalitis, hypoparathyroidism, pernicious anaemia.
 CC demyelinating diseases, cirrhosis, ulcerative colitis, dermatitis,
 CC myocarditis, regional enteritis, adult respiratory distress syndrome,
 CC infantile eczema, psoriasis lichen planus, allergic rhinitis, bronchial
 CC asthma, hypersensitivity, rheumatic fever and tissue rejection during
 CC transplantation
 XX
 SQ Sequence 395 AA;

Query Match 100.0%; Score 2057; DB 4; Length 395;
 Best Local Similarity 100.0%; Pred. No. 1.7e-205;
 Matches 395; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRLPRFSSTVMSLLMVQTGILVFLVSRQVPSPAGLGERVHVLLSSWRSGSFVGOLF 60
 DB 1 MRLPRFSSTVMSLLMVQTGILVFLVSRQVPSPAGLGERVHVLLSSWRSGSFVGOLF 60

QY 61 SQHPDVFYLMPEPAWHVWDTLSQGSAPALHMAVRDLIRSVFLCDMVDVFDAYLPWRRNISDL 120
 DB 61 SQHPDVFYLMPEPAWHVWDTLSQGSAPALHMAVRDLIRSVFLCDMVDVFDAYLPWRRNISDL 120

QY 121 FQWAVSRALCSPPVCEAFARGNISSEEVCKPLCATRPFGLAQACSSYSHVVLKEVRFN 180
 DB 121 FQWAVSRALCSPPVCEAFARGNISSEEVCKPLCATRPFGLAQACSSYSHVVLKEVRFN 180

QY 181 LOVLYPLSDPALNLRIVHLVRDPAVLRSDRQTKALARDNGIVLGTNGTWVEADPRLR 240
 DB 181 LOVLYPLSDPALNLRIVHLVRDPAVLRSDRQTKALARDNGIVLGTNGTWVEADPRLR 240

QY 241 VVNEVCRSHVRIAEALHKPPFPLODRYLRVRYEDLARDPLTVIRELYAFTGLGLTPQLQ 300
 DB 241 VVNEVCRSHVRIAEALHKPPFPLODRYLRVRYEDLARDPLTVIRELYAFTGLGLTPQLQ 300

QY 301 TWIHNITHGSGPGARREAFKTTSRDALSVSQAWRHHTLPFAKIRRVQELCGGALQLLGYRS 360
 DB 301 TWIHNITHGSGPGARREAFKTTSRDALSVSQAWRHHTLPFAKIRRVQELCGGALQLLGYRS 360

RESULT 2
 AAU11275
 ID AAU11275 standard; protein; 395 AA.
 XX
 AC AAU11275;
 XX
 DT 12-MAR-2002 (first entry)
 XX
 DE Murine intestinal-GlcNAc 6-sulfotransferase (I-GlcNAc6ST) protein.
 XX
 KW Mouse; beta1,3Gnt; beta1,3-N-acetylglucosaminyltransferase; MECA-79;
 KW L-selectin; L-selectin sulfotransferase-2; Crohn's disease; diabetes;
 KW ulcerative colitis; inflammatory skin disorder; psoriasis; Lichen planus;
 KW allergic contact dermatitis; lymphoma; chronic pneumonia; LSST-2;
 KW delayed-type hypersensitivity reaction; hyperplastic thymus; antiulcer;
 KW antiinflammatory; antipsoriatic; antidiabetic; dermatological;
 KW antiallergic; intestinal-GlcNAc 6-sulfotransferase; I-GlcNAc6ST.
 XX
 OS Mus musculus.
 XX
 PN WO200185177-A1.
 XX
 PD 15-NOV-2001.
 XX

PF 10-MAY-2001; 2001WO-US015452.
 XX
 PR 11-MAY-2000; 2000US-00569320.
 XX
 PA (BURN-) BURNHAM INST.
 XX
 PI Fukuda M, Yeh J, Hiraoka N;
 XX
 DR WPI; 2002-075226/10.
 DR N-PSDB; AAS16948.
 XX
 PT New enzyme, useful for modifying acceptor molecule, comprises an isolated
 PT L-selectin sulfotransferase-2 that directs expression of L-selectin
 PT ligand antigen, MECA-79 in Chinese hamster ovary cells, or intestinal
 PT GLCNAc 6-sulfotransferase.
 XX
 CC Claim 28; Fig 10; 98pp; English.
 XX
 CC The present invention provides a method of modifying an acceptor molecule
 CC by contacting the acceptor with an isolated beta1,3-N-
 CC acetylglucosaminyltransferase (beta1,3Gnt) or an active fragment, where
 CC beta1,3Gnt directs expression of a MECA-79 antigen. The invention also
 CC provides a method of treating or preventing an L-selectin-mediated
 CC condition by reducing the expression or activity of a beta1,3Gnt that
 CC directs expression of a MECA-79 antigen. This can be done by
 CC administering to the subject an oligosaccharide L-selectin antagonist
 CC that inhibits binding of L-selectin to a MECA-79 antigen, for example by
 CC administering antibody material that specifically binds beta1,3Gnt,
 CC and/or a beta1,3Gnt antisense nucleic acid molecule. L-selectin
 CC sulfotransferase-2 (LSST-2) also directs MECA-79 antigen expression.
 CC Alternatively, the expression or activity of LSST-2 or its active
 CC fragment can be reduced in combination with reducing the expression or
 CC activity of beta1,3Gnt. The method is useful for treating L-selectin
 CC mediated conditions such as Crohn's disease and ulcerative colitis,
 CC inflammatory disorders of the skin such as allergic contact dermatitis,
 CC psoriasis and Lichen planus, lymphomas, chronic pneumonia, delayed-type
 CC hypersensitivity reactions, diabetes and hyperplastic thymus. This
 CC sequence represents mouse I-GlcNAc6ST
 XX
 SQ Sequence 395 AA;

Query Match 100.0%; Score 2057; DB 5; Length 395;
 Best Local Similarity 100.0%; Pred. No. 1.7e-205;
 Matches 395; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRLPRFSSTVMSLLMVQTGILVFLVSRQVPSPAGLGERVHVLLSSWRSGSFVGOLF 60
 DB 1 MRLPRFSSTVMSLLMVQTGILVFLVSRQVPSPAGLGERVHVLLSSWRSGSFVGOLF 60

QY 61 SQHPDVFYLMPEPAWHVWDTLSQGSAPALHMAVRDLIRSVFLCDMVDVFDAYLPWRRNISDL 120
 DB 61 SQHPDVFYLMPEPAWHVWDTLSQGSAPALHMAVRDLIRSVFLCDMVDVFDAYLPWRRNISDL 120

QY 121 FQWAVSRALCSPPVCEAFARGNISSEEVCKPLCATRPFGLAQACSSYSHVVLKEVRFN 180
 DB 121 FQWAVSRALCSPPVCEAFARGNISSEEVCKPLCATRPFGLAQACSSYSHVVLKEVRFN 180

QY 181 LOVLYPLSDPALNLRIVHLVRDPAVLRSDRQTKALARDNGIVLGTNGTWVEADPRLR 240
 DB 181 LOVLYPLSDPALNLRIVHLVRDPAVLRSDRQTKALARDNGIVLGTNGTWVEADPRLR 240

QY 241 VVNEVCRSHVRIAEALHKPPFPLODRYLRVRYEDLARDPLTVIRELYAFTGLGLTPQLQ 300
 DB 241 VVNEVCRSHVRIAEALHKPPFPLODRYLRVRYEDLARDPLTVIRELYAFTGLGLTPQLQ 300

QY 301 TWIHNITHGSGPGARREAFKTTSRDALSVSQAWRHHTLPFAKIRRVQELCGGALQLLGYRS 360
 DB 301 TWIHNITHGSGPGARREAFKTTSRDALSVSQAWRHHTLPFAKIRRVQELCGGALQLLGYRS 360

361 VHSELEQRDLSLDLLPRGMDSFKNASSTKQPPES 395
 361 VHSELEQRDLSLDLLPRGMDSFKNASSTKQPPES 395

RESULT 3
ABB81557
ID ABB81557 standard; protein; 418 AA.
XX
AC ABB81557;
DT
DT 05-SEP-2002 (first entry)
XX
XX Mouse intestinal N-acetylglucosamine-6-sulfotransferase SEQ ID NO:5.
XX
XX Human; N-acetylglucosamine-6-sulfotransferase; enzyme; GlcNAc6ST;
KW corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
KW ophthalmological.
XX
OS Mus musculus.
XX
PN US2002061562-A1.
XX
XX 23-MAY-2002.
XX
XX 09-AUG-2001; 2001US-00927602.
XX
XX 11-AUG-2000; 2000US-00638211.
PR 11-AUG-2000; 2000US-0325773P.
XX
XX (FUKU/) FUKUDA M N.
PA (AKAM/) AKAMA T O.
XX
XX Fukuda MN, Akama TO;
PI
XX WPI; 2002-507643/54.
XX
XX New nucleic acid encoding corneal N-acetylglucosamine-6-sulfotransferase,
PT useful for treatment, monitoring and diagnosis of macular corneal
PT dystrophy.
XX
XX Example 5; Page 24-25; 69pp; English.
XX
XX The present invention describes human corneal N-acetylglucosamine-6-
CC sulfotransferase (I), which is able to catalyze sulfation of keratan
CC sulfate (KS). Also described is a method for monitoring the effect of
CC treatments for macular corneal dystrophy (MCD), and detecting
CC susceptibility to MCD. (I) is located to chromosome 16q22, and has
CC ophthalmological activity. (I) can be used to treat or prevent macular
CC corneal dystrophy types I or II. (I) makes possible treatment of MCD
CC without requiring keratoplasty or keratotomy. The present sequence
CC represents mouse intestinal N-acetylglucosamine-6- sulfotransferase,
CC which is given in comparison with (I) in the exemplification of the
CC present invention
XX
SQ Sequence 418 AA;

Query Match 99.3%; Score 2042.5; DB 5; Length 418;
Best Local Similarity 99.7%; Pred. No. 6e-204;
Matches 394; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 MRLPRFSTVMSLLMVTGTLVFLVSRQVPSAPGLGERVHVLVLSWRSGSFVGLP 60
Db 25 MRLPRFSTVMSLLMVTGTLVFLVSRQVPSAPGLGERVHVLVLSWRSGSFVGLP 84
Qy 61 SQHPDVFYLMPEPAHVMVDTLSQGSAPALHMAVRDLIRSVFLCDMDVFDAYLPWRNTSD 120
Db 85 SQHPDVFYLMPEPAHVMVDTLSQGSAPALHMAVRDLIRSVFLCDMDVFDAYLPWRNTSD 144
Qy 121 FQWAVSRALCSPVPCEAFARGNISSEEVCKPLCATRPFGLAQACSSYSHVVLKEVRF 180
Db 145 FQWAVSRALCSPVPCEAFARGNISSEEVCKPLCATRPFGLAQACSSYSHVVLKEVRF 204
Qy 181 LQVLYPLSDPALNLRIVHLVRDPAVLRSEQTAKALARDNGIVLGTNGTWVEADPRLR 240
Db 205 LQVLYPLSDPALNLRIVHLVRDPAVLRSEQTAKALARDNGIVLGTNGTWVEADPRLR 264

Qy 241 VNEVCRSHVRIAEAAHKKPPPPFLQDRYRLVRYEDLARDPLTVIRELYAFTGLTPOLO 300
Db 265 VNEVCRSHVRIAE-ALHKPPPPFLQDRYRLVRYEDLARDPLTVIRELYAFTGLTPOLO 323
Qy 301 TWIHNITHSGPGCARREAFKTTSRDALSVSQAWRHLPFAKIRRVQELCGGALQLLGYRS 360
Db 324 TWIHNITHSGPGCARREAFKTTSRDALSVSQAWRHLPFAKIRRVQELCGGALQLLGYRS 383
Qy 361 VHSLEQRDLSDLLPRGMDSFKWASSTEKQES 395
Db 384 VHSLEQRDLSDLLPRGMDSFKWASSTEKQES 418
RESULT 4
ABB81555
ID ABB81555 standard; protein; 395 AA.
XX
AC ABB81555;
XX
DT 05-SEP-2002 (first entry)
XX
XX Consensus N-acetylglucosamine-6-sulfotransferase SEQ ID NO:3.
XX
XX Human; N-acetylglucosamine-6-sulfotransferase; enzyme; GlcNAc6ST;
KW corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
KW ophthalmological.
XX
XX Homo sapiens.
OS Mus musculus.
OS Synthetic.
XX
XX Key Location/Qualifiers
FT Misc-difference 10 /label= Ala, Thr, Val
FT Misc-difference 13 /label= Ala, Val, Ser
FT Misc-difference 20 /label= Phe, Cys, Gly
FT Misc-difference 39 /label= Ala, Asp, Glu
FT Misc-difference 96 /label= Val, Met, Ile
FT Misc-difference 142 /label= Ala, Thr, Asn
FT Misc-difference 147 /label= Ala, Asp, Glu
FT Misc-difference 159 /label= Thr, Ser, Gly
FT Misc-difference 238 /label= Gly, His, Arg
FT Misc-difference 294 /label= Ser, Thr, Gly
FT Misc-difference 371 /label= Ala, Thr, Ser
FT Misc-difference 380 /label= Leu, Pro, Met
FT Misc-difference 382 /label= Gly, His, Ser
FT Misc-difference 384 /label= Thr, Ser, Lys
FT Misc-difference 390 /label= Ala, Glu
FT Misc-difference 391 /label= Ser, Lys
FT Misc-difference 392 /label= His, Gln
FT Misc-difference 394 /label= Arg, Glu
FT Misc-difference 395 /label= Asn, Ser
XX
XX US2002061562-A1.

PD 23-MAY-2002.
 XX
 XX 09-AUG-2001; 2001US-00927602.
 XX
 XX 11-AUG-2000; 2000US-00638211.
 PR 11-AUG-2000; 2000US-0325773P.
 XX
 PA (FUKU/) FUKUDA M N.
 PA (AKAM/) AKAMA T O.
 XX
 XX Fukuda MN, Akama TO;
 PI WPI; 2002-507643/54.
 XX
 XX New nucleic acid encoding corneal N-acetylglucosamine-6-sulfotransferase,
 PT useful for treatment, monitoring and diagnosis of macular corneal
 PT dystrophy.
 XX
 XX Example 5; Fig 2A-B; 69pp; English.
 XX
 CC The present invention describes human corneal N-acetylglucosamine-6-
 CC sulfotransferase (I), which is able to catalyze sulfation of keratan
 CC sulfate (KS). Also described is a method for monitoring the effect of
 CC treatments for macular corneal dystrophy (MCD), and detecting
 CC susceptibility to MCD. (I) is located to chromosome 16q22, and has
 CC ophthalmological activity. (I) can be used to treat or prevent macular
 CC corneal dystrophy types I or II. (I) makes possible treatment of MCD
 CC without requiring keratinoplasty or keratectomy. The present sequence
 CC represents a consensus N-acetylglucosamine-6-sulfotransferase which is
 CC given in the exemplification of the present invention
 XX
 XX Sequence 395 AA;
 SQ
 Query Match 84.0%; Score 1727; DB 5; Length 395;
 Best Local Similarity 84.7%; Pred. No. 5e-171;
 Matches 333; Conservative 16; Mismatches 44; Indels 0; Gaps 0;
 QY 1 MRLPRSTVMSLLMVQTLVLSRVQVPPSPAGLGERVHVLLVSSWRSGSFVGOLF 60
 DB 1 MMLPRFSSTVXTLLLAQTLLLVLSRPGSPAGGEXRHHVLLVSSWRSGSFVGOLF 60
 QY 61 SQHPDVFYLMPEPAWHVWDTLSQGSAPALHMAVRDLRSVFLCDMDVFDAYLPWRRNISDL 120
 DB 61 SQHPDVFYLMPEPAWHVWDTLSQGSAAATLHMAVRDLXSVFLCDMDVFDAYLPWRRNISDL 120
 QY 121 FQAVSRALCSPPVCEAFARGNISSEVEVKPLCATRPFGLAQACSSYSHVILKEVRFN 180
 DB 121 FQAVSRALCSPPACSAFPRGXISSEVCKPLCARQPFXLAREACRSYSHVILKEVRFN 180
 QY 181 LQVLYPLSDPALNLRIVHLVROPRAVLRSGEQTAKALARDNGVLGTNGTWVEADPRLR 240
 DB 181 LQVLYPLSDPALNLRIVHLVROPRAVLRSGEQTAKALARDNGVLGTNGTWVEADPXL 240
 QY 241 VVNEVCRSHVRIAEAAALHKPPPLQDRYLRVRYEDLARDPLTVIRELYAFTGLGLTPQLQ 300
 DB 241 VVREVCSSHVRIAEAAALTKPPPLRGYRLVRVEDLAREPLAEIRALYAFGLKLPQLE 300
 QY 301 TWIHNTHGSGPGARRAFKTTSDALSVSQAWRHLLTPFAKIRRVQELCGGALQLLGYSR 360
 DB 301 AWIHNTHGSGPGARRAFKTTSSRNALNVQAWRHLLTPFAKIRRVQELCGALQLLGYR 360
 QY 361 VHSLEORDLSLDLLPRGWDSEKWSSTEQP 393
 DB 361 VYSEDEQRDLXLDLVLPRGXDXFXWASSTXXXP 393
 RESULT 5
 ID AAY72640 standard; protein; 395 AA.
 XX
 AC AAY72640;
 XX
 DT 02-MAY-2001 (first entry)

XX Human glycosyl sulfotransferase-4beta (GST-4beta).
 DE
 XX Human; glycosyl sulfotransferase-4beta; GST-4beta; immunosuppressive;
 KW therapy; selectin binding inhibitor; gene therapy; inflammation;
 KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
 KW polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
 KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adenialitis;
 KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
 KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
 KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
 KW asthma; hypersensitivity; rheumatic fever; tissue rejection;
 KW chromosome 16q23.1.
 XX
 OS Homo sapiens.
 XX
 XX WO200106015-A1.
 PN
 XX 25-JAN-2001.
 PD
 XX 19-JUL-2000; 2000WO-US019741.
 PF
 XX 20-JUL-1999; 99US-0144694P.
 PR
 XX 13-JUN-2000; 2000US-00593828.
 PR
 XX (REGC) UNIV CALIFORNIA.
 PA
 XX Rosen SD, Lee JK, Hemmerich S;
 PI WPI; 2001-138471/14.
 DR
 XX N-PSDB; AAD02697, AAD02700.
 DR
 XX New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
 PT diagnostic and therapeutic agent screening applications.
 XX
 XX Claim 3; Fig 4B; 128pp; English.
 PS
 XX The present sequence is human glycosyl sulfotransferase-4beta (GST-4
 CC beta). GST-4 gene is found on chromosome 16q23.1. GST is a type 2
 CC membrane protein useful for inhibiting a binding event between a selectin
 CC and a selectin ligand, which comprises contacting the selectin with a non
 CC -sulphated selectin ligand, GST and a small molecular agent that inhibits
 CC the sulphation activity of GST. GST is also useful in inhibiting a
 CC selectin mediated binding event. GST is useful in gene therapy to treat
 CC disorders such as acute or chronic inflammation, systemic lupus
 CC erythematosus (SLE), rheumatoid arthritis, polyarteritis nodosa,
 CC polymyositis, dermatomyositis, systemic sclerosis, diabetes,
 CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
 CC disease, Grave's disease, adenialitis, hypoparathyroidism, pernicious
 CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
 CC dermatitis, myocarditis, regional enteritis, adult respiratory distress
 CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
 CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
 CC during transplantation
 XX
 XX Sequence 395 AA;
 SQ
 Query Match 83.0%; Score 1708; DB 4; Length 395;
 Best Local Similarity 83.0%; Pred. No. 4.8e-169;
 Matches 328; Conservative 24; Mismatches 43; Indels 0; Gaps 0;
 QY 1 MRLPRSTVMSLLMVQTLVLSRVQVPPSPAGLGERVHVLLVSSWRSGSFVGOLF 60
 DB 1 MMLPRVSSSTAVTALLAQITLLFLVSRPGSPAGGEARVHVLLVSSWRSGSFVGOLF 60
 QY 61 SQHPDVFYLMPEPAWHVWDTLSQGSAPALHMAVRDLRSVFLCDMDVFDAYLPWRRNISDL 120
 DB 61 NQHPDVFYLMPEPAWHVWDTLSQGSAAATLHMAVRDLVRSVFLCDMDVFDAYLPWRRNISDL 120
 QY 121 FQAVSRALCSPPVCEAFARGNISSEVEVKPLCATRPFGLAQACSSYSHVILKEVRFN 180
 DB 121 FQAVSRALCSPPACSAFPRGXISSEVCKPLCARQSFXLAREACRSYSHVILKEVRFN 180

Qy 181 LQVLYPLSDPALNLRIVHLVRDPRVLRSEOTAKALARDNGIVLGTNGTWVEADPRLR 240
Db 181 LQVLYPLSDPALNLRIVHLVRDPRVLRSEOTAKALARDNGIVLGTNGTWVEADPQLR 240
Qy 241 VVNEVCRSHVRIAEAAHLKPPFLQDRYLRVRYEDLARDPLTVIRELYAFTGLGLTPOLO 300
Db 241 VVNEVCRSHVRIAEAAHLKPPFLQDRYLRVRYEDLARDPLTVIRELYAFTGLGLTPOLE 300
Qy 301 TWIHNITHGSGPGARREAFKTTSDALSVSQAWRHLPFAKIRRVQELCGGALQLLGYRS 360
Db 301 TWIHNITHGSGPGARREAFKTTSDALSVSQAWRHLPFAKIRRVQELCGGALQLLGYRP 360
Qy 361 VHSELEQRDLSDLLPRGMDSPKWSSTEKQPS 395
Db 361 VYSEDEQRNLALDLVLPRLGNGFTWASSTASHPRN 395

RESULT 6
ABB81554
ID ABB81554 standard; protein; 395 AA.
XX
AC ABB81554;
XX
DT 05-SEP-2002 (first entry)
XX
DE Human corneal N-acetylglucosamine-6-sulfotransferase SEQ ID NO:2.
XX
KW Human; N-acetylglucosamine-6-sulfotransferase; enzyme; GlcNAc6ST;
KW corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
KW ophthalmological; chromosome 16q22.
XX
OS Homo sapiens.
XX
PN US2002061562-A1.
XX
PD 23-MAY-2002.
XX
PF 09-AUG-2001; 2001US-00927602.
XX
PR 11-AUG-2000; 2000US-00638211.
XX
PR 11-AUG-2000; 2000US-0325773P.
XX
PA (FUKU//) FUKUDA M N.
PA (AKAW//) AKAWA T O.
XX
XX Fukuda MN, Akama TO;
XX
XX WPI; 2002-507643/54.
DR N-PSDB; ABB89506.
XX
PT New nucleic acid encoding corneal N-acetylglucosamine-6-sulfotransferase,
PT useful for treatment, monitoring and diagnosis of macular corneal
PT dystrophy.
XX
PS Claim 13; Fig 1A-D; 69pp; English.
XX
CC The present sequence represents human corneal N-acetylglucosamine-6-
CC sulfotransferase (I), which is able to catalyse sulfation of keratan
CC sulfate (KS). Also described is a method for monitoring the effect of
CC treatments for macular corneal dystrophy (MCD), and detecting
CC susceptibility to MCD. (I) is located to chromosome 16q22, and has
CC ophthalmological activity. (I) can be used to treat or prevent macular
CC corneal dystrophy types I or II. (I) makes possible treatment of MCD
CC without requiring keratoplasty or keratotomy
XX
SQ Sequence 395 AA;

Query Match 83.0%; Score 1708; DB 5; Length 395;
Best Local Similarity 83.0%; Pred. No. 4.8e-169;
Matches 328; Conservative 24; Mismatches 43; Indels 0; Gaps 0;

Qy 1 MRLPRFSVTMLSLMVOTGILVFLVSRQVPSSPAGLGERVHVLVLSWSRSGSFVGQLF 60
Db 61 SQRPDPFYLMEPAWHVMDTILSQSGAPALHMAVRDLIRSVFLCDMDVFDAYLPMERNISDL 120
Db 61 NQHPDPFYLMEPAWHVMDTILSQSGAATLHMAVRDLVRSVFLCDMDVFDAYLPMERNISDL 120
Qy 121 FQWAVSRALCSPVCEAFARGNISSEVCKPLCATRPFGLAQEACSSYSHVVVKEVRFN 180
Db 121 FQWAVSRALCSPACSAFPRGAISSEAVCKPLCARQSFLLAREACRSYSHVVVKEVRFN 180
Qy 181 LQVLYPLSDPALNLRIVHLVRDPRVLRSEOTAKALARDNGIVLGTNGTWVEADPRLR 240
Db 181 LQVLYPLSDPALNLRIVHLVRDPRVLRSEOTAKALARDNGIVLGTNGTWVEADPQLR 240
Qy 241 VVNEVCRSHVRIAEAAHLKPPFLQDRYLRVRYEDLARDPLTVIRELYAFTGLGLTPOLO 300
Db 241 VVNEVCRSHVRIAEAAHLKPPFLQDRYLRVRYEDLAREPLAEIRALYAFGLSLTPQLE 300
Qy 301 TWIHNITHGSGPGARREAFKTTSDALSVSQAWRHLPFAKIRRVQELCGGALQLLGYRS 360
Db 301 TWIHNITHGSGPGARREAFKTTSDALSVSQAWRHLPFAKIRRVQELCGGALQLLGYRP 360
Qy 361 VHSELEQRDLSDLLPRGMDSPKWSSTEKQPS 395
Db 361 VYSEDEQRNLALDLVLPRLGNGFTWASSTASHPRN 395

RESULT 7
AAE15438
ID AAE15438 standard; protein; 395 AA.
XX
AC AAE15438;
XX
DT 12-MAR-2002 (first entry)
XX
DE Human drug metabolising enzyme (DME)-5.
XX
KW Human; drug metabolising enzyme; gene therapy; autoimmune disorder;
KW inflammatory disorder; acquired immune deficiency syndrome; infection;
KW AIDS; adult respiratory distress syndrome; cell proliferative disorder;
KW allergy; anaemia; conjunctivitis; actinic keratosis; arteriosclerosis;
KW cancer; endocrine disorder; hypothalamic disorder; pituitary disorder;
KW gastrointestinal disorder; metabolic disorder; developmental disorder;
KW liver disorder; iritis; cystic fibrosis; Addison's disease; retinitis;
KW goitre; diabetes; dysphagia; gastric carcinoma; anorexia; eye disorder;
KW DME-5.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..35 /label= Signal_peptide
FT Peptide 1..32 /label= Signal_peptide
FT Protein 33..395 /note= "Human mature DME-5 protein"
FT Protein 36..395 /note= "Human mature DME-5 protein"
XX
WO200179468-A2.
XX
PD 25-OCT-2001.
XX
PF 12-APR-2001; 2001WO-US011869.
XX
PR 13-APR-2000; 2000US-0197590P.
PR 19-APR-2000; 2000US-0198403P.
PR 28-APR-2000; 2000US-0200185P.
PR 05-MAY-2000; 2000US-020234P.
PR 11-MAY-2000; 2000US-0203509P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX

PI Policky JL, Hafalia A, Burford N, Ring HZ, Lal P, Tribouley CM;
 PI Yao MG, Yue H, Tang YT, Patterson C, Das D, Sanjanwala MS;
 PI Gandhi AR, Reddy R, Khan FA, Baughn MR, Ramkumar J, Griffin JA;
 PI Au-Young J;
 XX WPI; 2002-066363/09.
 DR N-PSDB; AAD24670.
 XX Novel isolated human drug metabolizing enzymes referred as DME 1-10,
 PT useful for diagnosing, treating, or preventing disorders associated with
 PT aberrant expression of DME such as allergy, anemia, asthma, infertility.
 XX Claim 1a; Page 131-132; 143pp; English.
 XX The invention relates to human drug metabolising enzymes referred as DME
 CC and nucleic acid molecules encoding such enzymes. Polynucleotides of the
 CC invention are useful for assessing toxicity of test compounds and in gene
 CC therapy. Sequences of the invention are useful in the diagnosis,
 CC prevention and treatment of autoimmune/inflammatory disorders such as
 CC acquired immune deficiency syndrome (AIDS), adult respiratory distress
 CC syndrome, allergies, anaemia, atherosclerosis, asthma, autoimmune
 CC haemolytic anaemia, contact dermatitis, Crohn's disease,
 CC glomerulonephritis, Goodpasture's syndrome, gout, Grave's disease,
 CC Hashimoto's thyroiditis, irritable bowel syndrome, multiple sclerosis,
 CC osteoarthritis, osteoporosis, psoriasis, systemic lupus erythematosus,
 CC rheumatoid arthritis, scleroderma, ulcerative colitis, uveitis, viral,
 CC bacterial, fungal, parasitic, protozoal, helminthic infections; cell
 CC proliferative disorders such as actinic keratosis, arteriosclerosis,
 CC atherosclerosis, Duchenne and Becker dystrophy, hepatitis, Cushing's
 CC syndrome, cancers, myelodysplastic syndrome, epilepsy; endocrine
 CC disorders such as disorders of the hypothalamus and pituitary resulting
 CC from lesions such as primary brain tumours, adenomas, infarction
 CC associated with pregnancy, aneurysms, vascular malformations; eye
 CC disorders such as conjunctivitis, iritis, retinitis, glaucoma, pigmentosa
 CC ; metabolic disorders such as Addison's disease, cystic fibrosis,
 CC diabetes, goitre, glycogen storage diseases, hypercholesterolaemia,
 CC hyperthyroidism, hypoglycaemia, lipid myopathies, Menkes syndrome,
 CC mannosidosis, obesity; gastrointestinal disorders such as dysphagia,
 CC gastric carcinoma, anorexia, nausea, gastroenteritis,
 CC hyperbilirubinaemia, emesis, cirrhosis, diarrhoea, jaundice, Reye's
 CC syndrome, peliosis hepatitis, hepatic vein thrombosis and developmental
 CC disorders. The present sequence is human DME-5 protein
 XX Sequence 395 AA;

Query Match 83.0%; Score 1708; DB 5; Length 395;
 Best Local Similarity 83.0%; Pred. No. 4.8e-169;
 Matches 328; Conservative 24; Mismatches 43; Indels 0; Gaps 0;

QY 1 MRLPFSSTVMSLLMVQGTGILVFLVSRQVPSPPAGIGERVHVILVSSWRSGSFGQLF 60
 DB 1 MWLPVUSSTAVTALLAQTFELLVFLVSRQVPSPPAGIGERVHVILVSSWRSGSFGQLF 60
 QY 61 SQHPDVFYLMPEAWHVDLTSQGSAPALHMAVRDLRSVFLCDMDVFDAYLPWRRNISDL 120
 DB 61 NQHPDVFYLMPEAWHVDLTSQGSATLHMAVRDLRSVFLCDMDVFDAYLPWRRNISDL 120
 QY 121 FQWAVSRALCSPVCEAFARGNISSEBVCXPLCNTRPFGLAQACSSYSHVLKEVRFN 180
 DB 121 FQWAVSRALCSPACSAFPRGAISSSEAVCKPLCARQSFLLAREACRSYSHVLKEVRFN 180
 QY 181 LQVLYPLLSPPALNLRVHVLPDRAVLRREQTAKALARDNGVLGTNGTWVEADPGLR 240
 DB 181 LQVLYPLLSPPALNLRVHVLPDRAVLRREQTAKALARDNGVLGTNGTWVEADPGLR 240
 QY 241 VVNEVCRSHVRIAEEAALHKPPPELQDRYLRVRYEDLARDPLTVIRELYAFTGLTLPQLQ 300
 DB 241 VVNEVCRSHVRIAEEAATLKPPPELGRYLRVRFEDLAREPLAEIRALYAFGLTLPQLQ 300
 QY 301 TWIHNTHGSGPGARRAFKTTSDALSVSQAWRHLLTPFAKIRRVQBLCGALQLLGYSR 360
 DB 301 AWIHNTHGSGPGARRAFKTTSRNALNVSAQRHALPFAKIRRVQBLCGALQLLGYSR 360

QY 361 VHSELEQRDLSDLLPRGMDSFKWASSTKOPES 395
 DB 361 VTSEDEQRNALDLVLPRGLNGFTWASSTASHPRN 395
 RESULT 8
 ADI21086
 ID ADI21086 standard; protein; 395 AA.
 XX ADI21086;
 AC ADI21086;
 XX 15-APR-2004 (first entry)
 DT
 XX Novel human protein #61.
 DE
 XX forensic; nutritional source; damaged tissue; diseased tissue;
 KW myeloid cell disorder; lymphoid cell disorder;
 KW bone cartilage tissue growth; tendon tissue growth;
 KW ligament tissue growth; nerve tissue growth; regeneration; wound healing;
 KW tissue repair; tissue replacement; burn; incision; ulcer; cancer; human.
 XX Homo sapiens.
 OS
 XX WO2003025148-A2.
 PN
 XX 27-MAR-2003.
 PD
 XX 19-SEP-2002; 2002WO-US029964.
 PF
 XX 19-SEP-2001; 2001US-0323739P.
 PR
 XX 13-SEP-2002; 2002US-00323739.
 XX (HYSE-) HYSEQ INC.
 PA
 XX Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;
 PI Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang D;
 PI Haley-Vicente D;
 XX WPI: 2003-354603/33.
 DR N-PSDB; ADI21802.
 XX New polynucleotides and secreted proteins, useful for treating myeloid or
 PT lymphoid cell disorders, in bone cartilage, tendon, ligament and nerve
 PT tissue growth or regeneration, in wound healing, and in tissue repair and
 PT replacement.
 XX Claim 20; SEQ ID NO 337; 156pp; English.

The invention relates to an isolated polynucleotide encoding a
 CC polypeptide with biological activity. The polynucleotides and
 CC polypeptides are useful in diagnostics, forensics, gene mapping,
 CC identification of mutations responsible for genetic disorders and other
 CC traits, to assess biodiversity, as nutritional sources or supplements.
 CC The polynucleotides may also be used as molecular weight markers,
 CC chromosome markers or map related gene positions, or as an antigen to
 CC raise anti-DNA antibodies or elicit immune response. The polypeptides are
 CC useful for raising antibodies, as markers for tissues in which the
 CC corresponding polypeptide is expressed, for re-engineering damaged or
 CC diseased tissues, for treating myeloid or lymphoid cell disorders, in
 CC bone cartilage, tendon, ligament and/or nerve tissue growth or
 CC regeneration, in wound healing, in tissue repair and replacement, in
 CC healing of burns, incisions and ulcers, and in treating cancer. The
 CC present sequence represents the amino acid sequence of a novel human
 CC protein.

XX Sequence 395 AA;

Query Match 83.0%; Score 1708; DB 7; Length 395;
 Best Local Similarity 83.0%; Pred. No. 4.8e-169;
 Matches 328; Conservative 24; Mismatches 43; Indels 0; Gaps 0;

QY 1 MRLPFSSTVMSLLMVQGTGILVFLVSRQVPSPPAGIGERVHVILVSSWRSGSFGQLF 60
 DB 1 MWLPVUSSTAVTALLAQTFELLVFLVSRQVPSPPAGIGERVHVILVSSWRSGSFGQLF 60

Db 1 MWLPRVSVSTAVTALLAQTFLLFLVSRPGPSSPAGGEARVHVLVLSWSRSGSFVQGLF 60
 QY 61 SQHPDVFYLMPEAWHVDWTLTSGSAPALHMAVRDLIRSVFLCDMDVDFDYLWRRNISDL 120
 Db 61 NQHPDVFYLMPEAWHVDWTLTSGSAPALHMAVRDLIRSVFLCDMDVDFDYLWRRNISDL 120
 QY 121 FQAVSRALCSPVPVCEAFARGNISSEVECKPLCATRPFGLAQEACSSYSHVVLKEVFFN 180
 Db 121 FQAVSRALCSPVPVCEAFARGNISSEVECKPLCATRPFGLAQEACSSYSHVVLKEVFFN 180
 QY 181 LQVLYPLSDPALNLRIVHLVRDPRVLRSEQTAKALARDNGIVLGTNGTWVEADPGLR 240
 Db 181 LQVLYPLSDPALNLRIVHLVRDPRVLRSEQTAKALARDNGIVLGTNGTWVEADPGLR 240
 QY 241 VVNEVCSSHVRIAEAAHKKPPFLQDRYLRVYEDLARDPLTVIRELYAFTGLGLTQ 300
 Db 241 VVNEVCSSHVRIAEAAHKKPPFLQDRYLRVYEDLARDPLTVIRELYAFTGLGLTQ 300
 QY 301 TWIHNTHGSGPGARREAFKTTSDALSVSQAWRHLPFAKIRRVQELCGALQLLYRS 360
 Db 301 TWIHNTHGSGPGARREAFKTTSDALSVSQAWRHLPFAKIRRVQELCGALQLLYRS 360
 QY 361 VHSLEQRDLSDLLPRGMDSFKWASSTKQPES 395
 Db 361 VYSEDEQRNLALDLVLRGLNGFTWASSTASHPRN 395

RESULT 9

ADL61235
 ID ADL61235 standard; protein; 395 AA.

XX AC ADL61235;
 XX XX
 DT 03-JUN-2004 (first entry)
 XX
 DE Human tyrosine kinase biomarker carbohydrate sulphotransferase 6 protein.
 XX
 KW predictor set; protein tyrosine kinase; cytostatic; antiangiogenic;
 KW vasotropic; vulnary; pharmacogenomic; drug sensitivity; breast cancer;
 KW hypervascular disease; angiogenesis; wound healing scar; human;
 KW biomarker; carbohydrate sulphotransferase 6; enzyme.
 XX

OS Homo sapiens.

XX WO2004020583-A2.

XX 11-MAR-2004.

XX 26-AUG-2003; 2003WO-US026491.

XX 27-AUG-2002; 2002US-0406385P.

XX (BRIM) BRISTOL-MYERS SQUIBB CO.

XX Huang F, Han X, Reeves KA, Amler L, Fairchild CR, Lee FY;

PI Shaw P;

XX WPI; 2004-239171/22.

DR N-PSDB; ADL61098.

XX New predictor sets with a plurality of polynucleotides and/or
 PT polypeptides whose expression pattern predicts cell response to a
 FT compound that modulates protein tyrosine kinase activity, useful in
 FT treating breast cancer.

XX Claim 9; SEQ ID NO 159; 649pp; English.

XX The invention relates to a novel predictor set comprising a plurality of
 CC polynucleotides and/or polypeptides whose expression pattern is
 CC predictive of the response of cells to treatment with a compound that
 CC modulates protein tyrosine kinase activity, or members of the protein
 CC tyrosine kinase pathway. The molecules of the invention demonstrate
 CC cytostatic, antiangiogenic, vasotropic and vulnary activities and may

CC be useful in the field of pharmacogenomics, in particular for determining
 .CC drug sensitivity and in treating breast cancer, hypervascular diseases,
 CC angiogenesis and scars in wound healing. The current sequence is that of
 CC a human protein tyrosine kinase biomarker protein of the invention.

XX Sequence 395 AA;

Query Match 83.0%; Score 1708; DB 8; Length 395;
 Best Local Similarity 83.0%; Pred. No. 4.8e-169;
 Matches 328; Conservative 24; Mismatches 43; Indels 0; Gaps 0;

QY 1 MRUPRSSTVMSLLMVQTCILVPLVSRQVPSAPGLGERVHVLVLSWSRSGSFVQGLF 60
 Db 1 MWLPRVSVSTAVTALLAQTFLLFLVSRPGPSSPAGGEARVHVLVLSWSRSGSFVQGLF 60
 QY 61 SQHPDVFYLMPEAWHVDWTLTSGSAPALHMAVRDLIRSVFLCDMDVDFDYLWRRNISDL 120
 Db 61 NQHPDVFYLMPEAWHVDWTLTSGSAPALHMAVRDLIRSVFLCDMDVDFDYLWRRNISDL 120
 QY 121 FQAVSRALCSPVPVCEAFARGNISSEVECKPLCATRPFGLAQEACSSYSHVVLKEVFFN 180
 Db 121 FQAVSRALCSPVPVCEAFARGNISSEVECKPLCATRPFGLAQEACSSYSHVVLKEVFFN 180
 QY 181 LQVLYPLSDPALNLRIVHLVRDPRVLRSEQTAKALARDNGIVLGTNGTWVEADPGLR 240
 Db 181 LQVLYPLSDPALNLRIVHLVRDPRVLRSEQTAKALARDNGIVLGTNGTWVEADPGLR 240
 QY 241 VVNEVCSSHVRIAEAAHKKPPFLQDRYLRVYEDLARDPLTVIRELYAFTGLGLTQ 300
 Db 241 VVNEVCSSHVRIAEAAHKKPPFLQDRYLRVYEDLARDPLTVIRELYAFTGLGLTQ 300
 QY 301 TWIHNTHGSGPGARREAFKTTSDALSVSQAWRHLPFAKIRRVQELCGALQLLYRS 360
 Db 301 TWIHNTHGSGPGARREAFKTTSDALSVSQAWRHLPFAKIRRVQELCGALQLLYRS 360
 QY 361 VHSLEQRDLSDLLPRGMDSFKWASSTKQPES 395
 Db 361 VYSEDEQRNLALDLVLRGLNGFTWASSTASHPRN 395

RESULT 10

AA72639

XX AA72639 standard; protein; 390 AA.

XX AC AA72639;

XX XX

DT 02-MAY-2001 (first entry)

DE Human glycosyl sulphotransferase-4alpha (GST-4alpha).

XX Human; glycosyl sulphotransferase-4alpha; GST-4alpha; immunosuppressive;
 KW therapy; selectin binding inhibitor; gene therapy; inflammation;
 KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
 KW polyarthritis nodosa; polymyositis; systemic sclerosis; dermatitis;
 KW Glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenalitis;
 KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
 KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
 KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
 KW asthma; hypersensitivity; rheumatic fever; tissue rejection;
 KW chromosome 16q23.1.

XX Homo sapiens.

OS XX

XX WO200106015-A1.

XX XX

XX 25-JAN-2001.

XX 19-JUL-2000; 2000WO-US019741.

XX 20-JUL-1999; 99US-0144694P.

PR 13-JUN-2000; 2000US-00593828.

XX (REGC) UNIV CALIFORNIA.

PA

XX Rosen SD, Lee JK, Hemmerich S;
XX WPI; 2001-138471/14.
XX N-PSDB; AAD02697, AAD02698, AAD02699.
XX
XX New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
XX diagnostic and therapeutic agent screening applications.
XX
XX Claim 3; Fig 1; 128pp; English.
XX
XX The present sequence is human glycosyl sulfotransferase-4alpha (GST-4
XX alpha). GST-4 gene is found on chromosome 16q23.1. GST is a type 2
XX membrane protein useful for inhibiting a binding event between a selectin
XX and a selectin ligand, which comprises contacting the selectin with a non
XX -sulphated selectin ligand, GST and a small molecular agent that inhibits
XX the sulphation activity of GST. GST is also useful in inhibiting a
XX selectin mediated binding event. GST is useful in gene therapy to treat
XX disorders such as acute or chronic inflammation, systemic lupus
XX erythematosus (SLE), rheumatoid arthritis, polyarteritis nodosa,
XX polymyositis, dermatomyositis, systemic sclerosis, diabetes,
XX glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
XX disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious
XX anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
XX dermatitis, myocarditis, regional enteritis, adult respiratory distress
XX syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
XX bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
XX during transplantation. Note: The present sequence is also shown in
XX sequence listing (page no: 56) but lacks four nucleotides at its 3' end
XX
XX Sequence 390 AA;

Query Match 74.7%; Score 1536.5; DB 4; Length 390;
Best Local Similarity 76.0%; Pred. No. 3.9e-151;
Matches 295; Conservative 30; Mismatches 62; Indels 1; Gaps 1;
QY 1 MRLPRFSSVTMLSLMLVQGTG-ILVFLVSRQVSPSPAGLGERVHVLVLSWRSGSSFFVQGL 59
Db 1 MWLPRFSSKTVTVLLLAQTCLLLFIISRPCSPSPAGGEDRVHVLVLSWRSGSSFFLQGL 60
QY 60 FSQHPDVFYLMPEPAHWVDTLSQGSAPALHMAVRDLIRSVFLCDMDVDFDAYLPMWRNID 119
Db 61 FSQHPDVFYLMPEPAHWVDTLSQGSAAATLHMAVRDLIRSVFLCDMDVDFDAYLPMQSRNLSA 120
QY 120 LFQWNSRALCSPVCEAFARGNISSEVECKPLCATPFGLAQACSSYSHVVLKEVRF 179
Db 121 FFWNATSRALCSPPACSAFPRGTISKQDVCKTCTQFPFLAREACRSYSHVVLKEVRF 180
QY 180 NLQVLYPLLSDDPALNLRIVHLVRDPRAVLRSDREQTAKALARDNGIVLGTNGTWVEADPRL 239
Db 181 NLQVLYPLLSDDPALNLRIVHLVRDPRAVLRSDREAGPILARDNGIVLGTNGKWEADPHL 240
QY 240 RVNVECRSHVRIAEAAALHKPPPPFLQDRYRLRYVEDIARDPLTVIRELYAFTGLGLTPQL 299
Db 241 RLIREVCRSHVRIAEAAATLKPFPFLGRYRLRVFEDLAREPLAEIRALYAFTGLGLTPQL 300
QY 300 QTWIHNTHGSGPGARREAFKTTSDALSVSQAWRHLPFAKIRRVQELCGGALQLLYR 359
Db 301 EAWIHNTHGSGIGKPIEAFHTSSRNARNVSQAWRHLPFTKILRVQECAGALQLLYR 360
QY 360 SVHSELEQRDLSDLLPRGMDSPFKWAS 387
Db 361 PVISADQQRDLTLVLPRGPDHESWAS 388

RESULT 11
ABB81556
ID ABB81556 standard; protein; 390 AA.
XX
AC ABB81556;
XX
DT 05-SEP-2002 (first entry)
XX

DE Human intestinal N-acetylglucosamine-6-sulfotransferase SEQ ID NO:4.
XX
XX Human; N-acetylglucosamine-6-sulfotransferase; enzyme; GlcNAc6ST;
XX corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
XX ophthalmological.
XX
XX Homo sapiens.
XX
XX US2002061562-A1.
XX
XX 23-MAY-2002.
XX
XX 09-AUG-2001; 2001US-00927602.
XX
XX 11-AUG-2000; 2000US-00638211.
XX
XX 11-AUG-2000; 2000US-0325773P.
XX
XX (FUKU/) FUKUDA M N.
XX (AKAM/) AKAMA T O.
XX
XX Fukuda MN, Akama TO;
XX
XX WPI; 2002-507643/54.
XX
XX New nucleic acid encoding corneal N-acetylglucosamine-6-sulfotransferase,
XX useful for treatment, monitoring and diagnosis of macular corneal
XX dystrophy.
XX
XX Example 5; Fig 2A-B; 69pp; English.
XX
XX The present invention describes human corneal N-acetylglucosamine-6-
XX sulfotransferase (I), which is able to catalyze sulfation of keratan
XX sulfate (KS). Also described is a method for monitoring the effect of
XX treatments for macular corneal dystrophy (MCD), and detecting a
XX susceptibility to MCD. (I) is located to chromosome 16q22, and has
XX ophthalmological activity. (I) can be used to treat or prevent macular
XX corneal dystrophy types I or II. (I) makes possible treatment of MCD
XX without requiring keratoplasty or keratectomy. The present sequence
XX represents human intestinal N-acetylglucosamine-6- sulfotransferase,
XX which is given in comparison with (I) in the exemplification of the
XX present invention
XX
XX Sequence 390 AA;

Query Match 74.7%; Score 1536.5; DB 5; Length 390;
Best Local Similarity 76.0%; Pred. No. 3.9e-151;
Matches 295; Conservative 30; Mismatches 62; Indels 1; Gaps 1;
QY 1 MRLPRFSSVTMLSLMLVQGTG-ILVFLVSRQVSPSPAGLGERVHVLVLSWRSGSSFFVQGL 59
Db 1 MWLPRFSSKTVTVLLLAQTCLLLFIISRPCSPSPAGGEDRVHVLVLSWRSGSSFFLQGL 60
QY 60 FSQHPDVFYLMPEPAHWVDTLSQGSAPALHMAVRDLIRSVFLCDMDVDFDAYLPMWRNID 119
Db 61 FSQHPDVFYLMPEPAHWVDTLSQGSAAATLHMAVRDLIRSVFLCDMDVDFDAYLPMQSRNLSA 120
QY 120 LFQWNSRALCSPVCEAFARGNISSEVECKPLCATPFGLAQACSSYSHVVLKEVRF 179
Db 121 FFWNATSRALCSPPACSAFPRGTISKQDVCKTCTQFPFLAREACRSYSHVVLKEVRF 180
QY 180 NLQVLYPLLSDDPALNLRIVHLVRDPRAVLRSDREQTAKALARDNGIVLGTNGTWVEADPRL 239
Db 181 NLQVLYPLLSDDPALNLRIVHLVRDPRAVLRSDREAGPILARDNGIVLGTNGKWEADPHL 240
QY 240 RVNVECRSHVRIAEAAALHKPPPPFLQDRYRLRYVEDIARDPLTVIRELYAFTGLGLTPQL 299
Db 241 RLIREVCRSHVRIAEAAATLKPFPFLGRYRLRVFEDLAREPLAEIRALYAFTGLGLTPQL 300
QY 300 QTWIHNTHGSGPGARREAFKTTSDALSVSQAWRHLPFAKIRRVQELCGGALQLLYR 359
Db 301 EAWIHNTHGSGIGKPIEAFHTSSRNARNVSQAWRHLPFTKILRVQECAGALQLLYR 360
QY 360 SVHSELEQRDLSDLLPRGMDSPFKWAS 387

Db 361 PVSAQQQRLTLDLVLPGRDPHFSWAS 388
 ||| :|||:||||:||||| |||
 RESULT 12
 AAB41947
 ID AAB41947 standard; protein; 418 AA.
 AC AAB41947;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE Human ORFX ORF1711 polypeptide sequence SEQ ID NO:3422.
 XX
 KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnary; antiparasitic; antiparkinsonian; neurotropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antineoplastic;
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive.
 XX
 OS Homo sapiens.
 XX
 XX WO200058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 XX 31-MAR-2000; 2000WO-US008621.
 XX
 PR 31-MAR-1999; 99US-0127607P.
 PR 02-APR-1999; 99US-0127636P.
 PR 05-APR-1999; 99US-0127728P.
 PR 30-MAR-2000; 2000US-00540763.
 XX
 XX (CURA-) CURAGEN CORP.
 XX
 XX Shimkets RA, Leach M;
 XX
 DR WPI; 2000-602362/57.
 DR N-PSDB; AAC76156.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease.
 XX
 PS Claim 11; Page 2599-2600; 5507pp; English.
 XX
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
 CC antiparasitic; antiparkinsonian; neurotropic; neuroprotective; osteopathic;
 CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
 CC cardiac; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
 CC dermatological; immunosuppressive; antineoplastic; antithyroid;
 CC antiviral; antifungal; antirheumatic; antithyroid; and antianaemic. The
 CC sequences can be used for determining the presence of or predisposition
 CC to, or preventing or treating pathological conditions associated with an
 CC ORFX-associated disorder. The nucleic acids can be used to express ORFX
 CC proteins in gene therapy vectors. The proteins and nucleic acids may be
 CC used to treat cancers, proliferative disorders, neurodegenerative
 CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
 CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
 CC storage, systemic lupus erythematosus, severe combined immunodeficiency
 CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
 CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and

CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
 CC enhance coagulation; to inhibit thrombosis; and as a contraceptive
 XX
 SQ Sequence 418 AA;
 Query Match 69.8%; Score 1436; DB 3; Length 418;
 Best Local Similarity 72.0%; Pred. No. 1.4e-140;
 Matches 280; Conservative 34; Mismatches 73; Indels 2; Gaps 2;
 QY 1 MRLPFSSTVMSLLMVQTG-ILVFLVSRQVSPSPAGLGERVHVLVLSWSRSGSFGQL 59
 DB 28 MWLPFSSKTIVLLLAQTCLLFIISRPSPGAGEDRVHVLVLSWSRSGSFGQL 87
 QY 60 FSQHPDVYLMPEAHVMDTLSSQSPALHMAVRDLIRSVFLCDMDVFDAYL-PWRNIS 118
 DB 88 FSQHPDVYLMPEAHVMDTLSSQSAATLHMAVRDLIRSVFLCDMDVFDAYMEGPRRS 147
 QY 119 DLQWAVSRALCPFPCEAFARGNISSEVECKPLCATRPFGLAQEACSSYSHVVLKEVRF 178
 DB 148 SLQWENSRLCAPACDIIPQDEIIPRAHCRLLCSQPPFEVVEKACRSYSHVVLKEVRF 207
 QY 179 FNLQVLYPLLSDPALNLRIVHLVRDPAVLRSEOTAKALARDNGIVLGTNGTWVEADPR 238
 DB 208 FNLQVLYPLLSKPSLNHIVHLVRDPAVLRSEAGPILARDNGIVLGTNGKWEADPH 267
 QY 239 LRVVNEVCRSHVRIAEAAHLKPPFPFLQDRLVRYEDLARDPLTVIRELYAFTGLTPQ 298
 DB 268 LRLIREVCRSHVRIAEAAHLKPPFPFLGRVRLVRFEDLAREPLAIRALYFTGLTPQ 327
 QY 299 LQWVHNTGSGPGARRBAFKTTSRDALSVSQNRHTLPAKIRRVQELCGALQLGY 358
 DB 328 LEAWIHNTGSGIGKPIEAFHTSSRNARVNSQNRHALPFTKILRVQEVCAQALQLGY 387
 QY 359 RSVHSELEQRDLSLDLLPRGMDSEKWS 387
 DB 388 RPYVSAQQQRLTLDLVLPGRDPHFSWAS 416
 RESULT 13
 ADJ70405
 ID ADJ70405 standard; protein; 394 AA.
 XX
 AC ADJ70405;
 XX
 DT 06-MAY-2004 (first entry)
 XX
 DE Human heat mitochondrial protein as a therapeutic target SeqID2211.
 KW mitochondrial; human; screening assay; diabetes mellitus;
 KW Huntington's disease; osteoarthritis;
 KW Leber's hereditary optic neuropathy; LHON;
 KW mitochondrial encephalopathy lactic acidosis and stroke; MELAS;
 KW myoclonic epilepsy ragged red fibre syndrome; MERRF; cancer;
 KW neuroprotective; neurotropic; antidiabetic; anticonvulsant; antiarthritic;
 KW osteopathic; ophthalmological; cytostatic.
 XX
 OS Homo sapiens.
 XX
 XX WO2003087768-A2.
 XX
 PD 23-OCT-2003.
 XX
 PF 04-APR-2003; 2003WO-US010870.
 XX
 PR 12-APR-2002; 2002US-0372843P.
 PR 17-JUN-2002; 2002US-0389987P.
 PR 20-SEP-2002; 2002US-0412418P.
 XX
 PA (MITO-) MITOKOR.
 PA (BUCK-) BUCK INST AGE RES.
 XX
 XX Ghosh SS, Fahy ED, Zhang B, Gibson BW, Taylor SW, Glenn GM;
 PI Warnock DE;

XX WPI; 2003-845369/78.
 XX Identifying a mitochondrial target for drug screening assays and for
 PT treating diseases associated with altered mitochondrial function,
 PT comprises detecting a modified polypeptide in a sample and correlating
 PT with the disease.
 XX
 XX Claim 1; SEQ ID NO 2211; 180pp; English.
 XX
 CC This invention relates to novel mitochondrial targets that can be used
 CC for therapeutic intervention in treating a disease associated with
 CC altered mitochondrial function. Specifically, it refers to a method for
 CC identifying proteins of the human heart mitochondrial proteome that are
 CC useful for drug screening assays, as well as therapeutic targets. The
 CC present invention describes a method for identifying such proteins that
 CC can be used in the treatment of various diseases associated with altered
 CC mitochondrial function including diabetes mellitus, Huntington's disease,
 CC osteoarthritis, Leber's hereditary optic neuropathy (LHON), mitochondrial
 CC encephalopathy lactic acidosis and stroke (MELAS), myoclonic epilepsy
 CC ragged red fibre syndrome (MERRF) or cancer. Accordingly, these
 CC compositions have neuroprotective, neurotropic, antidiabetic,
 CC anticonvulsant, antiarthritic, osteopathic, ophthalmological, and
 CC cytotatic activities. This polypeptide sequence is a human heart
 CC mitochondrial protein of the invention.
 XX
 XX Sequence 394 AA;

Query Match 64.8%; Score 1333; DB 7; Length 394;
 Best Local Similarity 78.1%; Pred. No. 7.2e-130;
 Matches 257; Conservative 19; Mismatches 47; Indels 6; Gaps 1;

QY 1 MRLPRSSVTMLSLMVOTGILVPLVSQVPSPPAGLGERVHVLVLSMWSGSGFVGOLF 60
 DB 1 MWLPRVSTAVTALLAQTFLLVLSVPGSPAGGEARVHVLVLSMWSGSGFVGOLF 60
 QY 61 SQHPDVFYLMPEAWHVMWDTLSQGSAPALHMAVRDLIRSVFLCDMDVFDAYLPWRRNISDL 120
 DB 61 NQHPDVFYLMPEAWHVMWDTLSQGSAAALHMAVRDLIRSVFLCDMDVFDAYLPWRRNISDL 120
 QY 121 FQWAVSRALCSPVCEAFARGNISSEVCKPLCATRPFGLAQACSSYSHVVLKEVRFEN 180
 DB 121 FQWAVSRALCSPVCEAFARGNISSEVCKPLCATRPFGLAQACSSYSHVVLKEVRFEN 180
 QY 181 LOVLYPLLSDPALNLRIVHLVRDPRAVLRSEQTAKALARDNGIVLGTNGTWVEADPRLR 240
 DB 181 LOVLYPLLSDPALNLRIVHLVRDPRAVLRSEQTAKALARDNGIVLGTNGTWVEADPRLR 240
 QY 241 VVNEVCRSHVRIAEAAALHKPPPLQDRYLRVRYEDLARDPLTVIRELYAFTGLTLPQLQ 300
 DB 241 LIREVCRSHVRIAEAAALHKPPPLQDRYLRVRYEDLARDPLTVIRELYAFTGLTLPQLQ 300
 QY 301 TWIHNTHGSG-----PGARRRAFKTTS 323
 DB 301 AWIHNTHGSGASQSKPSILRLGMRATS 329

RESULT 14
 AAM93309
 ID AAM93309 standard; protein; 386 AA.
 XX
 AC AAM93309;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human polypeptide, SEQ ID NO: 2817.
 XX
 KW Human; full length cDNA; cDNA synthesis; oligo-capping.
 XX
 OS Homo sapiens.
 XX
 XX EPI130094-A2.
 XX

PD 05-SEP-2001.
 XX
 PF 07-JUL-2000; 2000EP-00114089.
 XX
 PR 08-JUL-1999; 99JP-00194486.
 PR 11-JAN-2000; 2000JP-00118774.
 PR 02-MAY-2000; 2000JP-00183765.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 XX
 DR WPI; 2001-524255/58.
 DR N-PSDB; AAK94229.
 XX
 XX 830 Primers useful for synthesizing full length cDNA clones and their use
 PT in genetic manipulation.
 PT
 PS Claim 8; SEQ ID NO:2817; 1380pp + Sequence Listing; English.
 XX
 CC The invention relates to primers for synthesising full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been isolated
 CC and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have
 CC been determined. Primers for synthesising the full length cDNA are useful
 CC for clarifying the function of the protein encoded by the cDNA. The full
 CC length clones were obtained by construction of full length enriched cDNA
 CC libraries that were synthesised by the oligo-capping method. The primers
 CC enable the production of the full length cDNA easily without any special
 CC methods. The present sequence is a polypeptide encoded by a full length
 CC human cDNA of the invention. Note: The sequence data for this patent did
 CC not form part of the printed specification, but was obtained in CD-ROM
 CC format directly from EPO
 XX
 XX Sequence 386 AA;
 SQ
 Query Match 49.7%; Score 1022.5; DB 4; Length 386;
 Best Local Similarity 56.7%; Pred. No. 1.9e-97;
 Matches 216; Conservative 42; Mismatches 112; Indels 11; Gaps 7;
 QY 1 MRLPRSSVTMLSLMVOTGILVPLVSQVPSPPAGLGERVHVLVLSMWSGSGSFV 56
 DB 1 MLLPK--KKMLLLFLVSQMAILALFFHMYSHNLSLSMKAKAQPMMHVLVLSMWSGSGSFV 58
 QY 57 GOLFSQHPDVFYLMPEAWHVMWDTLSQGSAPALHMAVRDLIRSVFLCDMDVFDAYLPWRR 115
 DB 59 GQLFGQHPDVFYLMPEAWHVMWDTLSQGSAPALHMAVRDLIRSVFLCDMDVFDAYLPWRR 118
 QY 116 NISDLFQWAVSRALCSPVCEAFARGNISSEVCKPLCATRPFGLAQACSSYSHVVLKE 175
 DB 119 RQSSSLFQWENSALCSPVCEAFARGNISSEVCKPLCATRPFGLAQACSSYSHVVLKE 178
 QY 176 VRFFNLQVLYPLLSDPALNLRIVHLVRDPRAVLRSEQTAKALARDNGIVLGTNGTWV-E 234
 DB 179 VRFFNLQVLYPLLSDPALNLRIVHLVRDPRAVLRSEQTAKALARDNGIVLGTNGTWV-E 238
 QY 235 ADPLRVLVNEVCRSHVRIAEAAALHKPPPLQDRYLRVRYEDLARDPLTVIRELYAFTGLG 294
 DB 239 EDQPYVMQVICSQSEIYK-TIQSLPKALQERLLVRYEDLARDPLTVIRELYAFTGLG 297
 QY 295 LTPQLQTIWNIHNGSGPGARREAFKTSRDALSVSQAWRHHTLPFAKIRRVQELCGGALQ 354
 DB 298 FLPHLQTIWNIHNGSGPGARREAFKTSRDALSVSQAWRHHTLPFAKIRRVQELCGGALQ 355
 QY 355 LLGYRSHVSELEQORDLSLDLL 375
 DB 356 LLGYRSHVSELEQORDLSLDLL 376
 RESULT 15
 ADL30784
 ID ADL30784 standard; protein; 386 AA.
 XX

AC ADL30784;
XX 20-MAY-2004 (first entry)
XX Human protein encoded by a full length cDNA clone SeqID 2817.
DE human; medicine; signal transduction; glycoprotein; transcription;
KW oligo-capping method.
XX Homo sapiens.
XX EPI396543-A2.
XX 10-MAR-2004.
XX 07-JUL-2000; 2003EP-00025638.
XX 08-JUL-1999; 99JP-00194486.
PR 11-JAN-2000; 2000JP-00118774.
PR 02-MAY-2000; 2000JP-00183865.
PR 07-JUL-2000; 2000EP-00114089.
XX (REAS-) RES ASSOC BIOTECHNOLOGY.
XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX WPI; 2004-204755/20.
DR N-PSDB; ADL30783.
XX New oligonucleotide primers (830 cDNAs) useful for synthesizing full
PT length human cDNAs.
XX Example 1; SEQ ID NO 2817; 1340pp; English.
XX This invention relates to a novel primers useful for synthesising full
CC length cDNA molecules that encode human proteins. Specifically, it refers
CC to secretory or membrane proteins that are potential therapeutic agents/
CC target molecules in the field of medicine, and in particular genes
CC encoding proteins that are associated with signal transduction,
CC glycoproteins and transcription. The present invention describes a method
CC for efficiently cloning a full length human cDNA from both the 5' and 3',
CC ends using the oligo-capping method. This polypeptide sequence is a full
CC length human protein of the invention.
XX Sequence 386 AA;
SQ Query Match 49.7%; Score 1022.5; DB 8; Length 386;
Best Local Similarity 56.7%; Pred. No. 1.9e-97;
Matches 216; Conservative 42; Mismatches 112; Indels 11; Gaps 7;
Qy 1 MRLPRFSTVMSLLMYQTGILVP---LVSRQVPS-SPAGLGERVHVHVLSSWRSGSFV 56
Db 1 MLLPK--KMKLLFLVSQMAILALFFHMYSHNLSLSMKAPERMHVHVLSSWRSGSFV 58
Qy 57 GOLFSQHPDVFYLMPEAHVWVDTLSCGSAPALHMAVRDLRSVFLCDMDVFDAYL-PWRR 115
Db 59 GOLFGQHPDVFYLMPEAHVWVDTLSCGSAPALHMAVRDLRSVFLCDMSVFDAYMEGPR 118
Qy 116 NISDLQWAVSRALCSPVCEAFARGNISSEVCKPLCATRPFGLAQEACSSYSHVVLKE 175
Db 119 QSSSLFQWENSALCSPACDIIIPQDEIIIPRAHCLLCQQPFVEVEKACRSYSHVVLKE 178
Qy 176 VRFNQLVLYPLSDPALNLRIVHLVDRDPRAVLSRSQTAKALARDNGIVLGTNGTWV-E 234
Db 179 VRFNQLVLYPLSDPALNLRIVHLVDRDPRAVLSRSQTAKALARDNGIVLGTNGTWV-E 238
Qy 235 ADPLRVNVECRSHVRAEAAHKKPPFPLODRVLRVYEDLARDPLTVIRELYAFTGLG 294
Db 239 EDQPYVNVQVTCQSOLEIYK-TIOSLPKALQERYLLVRYEDLARAPVAQTSRMTEFVGL 297
Qy 295 LTPOLQWVHNHITHGSGPGARRRPAKTTSDALSQAWRHHTLPFAKIRRVQELCGGALQ 354

Db 298 FLPHLOTWVHNITRGKMG--DHAFTNARDALNVSQAWRWSLPYEKVSRQKACGDAMN 355
Qy 355 LLGYRSVHSELEQORDLSLDLL 375
Db 356 LLGYRHRVSRSEQQRNLLLDLL 376
Search completed: June 23, 2005, 08:43:22
Job time : 45.6885 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2005, 08:33:38 ; Search time 8.95931 Seconds
(without alignments)
4242.027 Million cell updates/sec

Title: US-10-697-828-7

Perfect score: 2057

Sequence: 1 MRLPRFSSTVMSLLMVQTG.....LPRGMSFKWASSTKQPS 395

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 79:*

1: Pirl1:*

2: Pirl2:*

3: Pirl3:*

4: Pirl4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | ID | Description |
|------------|-------|-------|--------|----------|--------------------|
| 1 | 643.5 | 31.3 | 484 | 2 JE0261 | N-acetylglucosamin |
| 2 | 580 | 28.2 | 486 | 2 JC7351 | N-acetylglucosamin |
| 3 | 577 | 28.1 | 484 | 2 JC7350 | N-acetylglucosamin |
| 4 | 495 | 24.1 | 458 | 2 A57397 | chondroitin 6-sulf |
| 5 | 148.5 | 7.2 | 307 | 2 E95934 | probable enzyme, C |
| 6 | 126 | 6.1 | 388 | 2 G70729 | hypothetical prote |
| 7 | 101.5 | 4.9 | 262 | 2 T35999 | probable aminoglyc |
| 8 | 99 | 4.8 | 571 | 2 AG3174 | acetolactate synth |
| 9 | 98 | 4.8 | 856 | 2 F87316 | conserved hypothet |
| 10 | 95 | 4.6 | 547 | 2 A64878 | peptide transport |
| 11 | 94.5 | 4.6 | 892 | 1 S42228 | replication licens |
| 12 | 93 | 4.5 | 547 | 2 G90862 | hypothetical prote |
| 13 | 93 | 4.5 | 547 | 2 B85756 | hypothetical prote |
| 14 | 93 | 4.5 | 559 | 2 T03412 | malate synthase (E |
| 15 | 92.5 | 4.5 | 327 | 2 T50744 | spheroidene monoox |
| 16 | 92.5 | 4.5 | 4077 | 2 T17484 | hypothetical prote |
| 17 | 91 | 4.4 | 562 | 2 T48413 | malate synthase-li |
| 18 | 91 | 4.4 | 925 | 2 T02811 | DNA excision/repai |
| 19 | 90 | 4.4 | 933 | 1 DEBOG | oxoglutarate dehyd |
| 20 | 90 | 4.4 | 933 | 2 F85573 | hypothetical prote |
| 21 | 90 | 4.4 | 933 | 2 G90722 | oxoglutarate dehyd |
| 22 | 89.5 | 4.4 | 311 | 2 B40216 | flavonol 3'-sulfol |
| 23 | 89.5 | 4.4 | 459 | 2 G83174 | exodeoxyribonuclea |
| 24 | 88 | 4.3 | 668 | 2 T01685 | crpl protein - mai |
| 25 | 87.5 | 4.3 | 342 | 2 T49687 | lipoyltransferase |
| 26 | 87.5 | 4.3 | 678 | 2 T23341 | beta-catenin - Cae |
| 27 | 87.5 | 4.3 | 860 | 2 C82750 | mannosyltransferas |
| 28 | 87.5 | 4.3 | 1158 | 2 F90854 | probable host spec |
| 29 | 87 | 4.2 | 413 | 2 H86825 | hypothetical prote |

| | | | | | |
|----|------|-----|------|----------|--------------------|
| 30 | 87 | 4.2 | 436 | 2 E90261 | hypothetical prote |
| 31 | 87 | 4.2 | 643 | 2 B72602 | hypothetical prote |
| 32 | 87 | 4.2 | 814 | 2 B75050 | secretory protein |
| 33 | 87 | 4.2 | 1118 | 2 C95385 | probable adenylate |
| 34 | 86.5 | 4.2 | 301 | 2 D87684 | transcription regu |
| 35 | 86.5 | 4.2 | 327 | 2 S49819 | crta protein - Rho |
| 36 | 86.5 | 4.2 | 899 | 2 B38529 | nikB protein - Esc |
| 37 | 86.5 | 4.2 | 1092 | 2 H69071 | DNA-directed DNA p |
| 38 | 86 | 4.2 | 652 | 2 T45499 | probable DNA recom |
| 39 | 86 | 4.2 | 950 | 2 T38777 | probable valyl trn |
| 40 | 86 | 4.2 | 1165 | 2 D85842 | probable tail fibe |
| 41 | 86 | 4.2 | 1185 | 2 F90877 | probable host spec |
| 42 | 86 | 4.2 | 3796 | 2 T18514 | lysosomal traffick |
| 43 | 85.5 | 4.2 | 385 | 2 S72702 | 8-amino-7-oxononan |
| 44 | 85.5 | 4.2 | 425 | 2 E83023 | 3-deoxy-D-manno-oc |
| 45 | 85.5 | 4.2 | 532 | 2 S33758 | flavin-containing |

ALIGNMENTS

RESULT 1

JE0261

N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) - human

C/Species: Homo sapiens (man)

C/Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 01-Mar-2002

C/Accession: JE0261

R/Uchimura, K.; Muramatsu, H.; Kaname, T.; Ogawa, H.; Yamakawa, T.; Fan, Q.W.; Mitsuoka, T.

J. Biochem. 124, 670-678, 1998

A/Title: Human N-Acetylglucosamine-6-O-sulfotransferase involved in the biosynthesis of (

A/Reference number: JE0261; MUID:98391845; PMID:9722682

A/Accession: JE0261

A/Molecule type: mRNA

A/Residues: 1-484 <UCH>

A/Cross-references: DDBJ:AB014679

C/Comment: This protein catalyzes the transfer of sulfur of sulfate from 3'-phosphoadenos

C/Superfamily: chondroitin 6-sulfotransferase

C/Keywords: sulfotransferase

Query Match 31.3%; Score 643.5; DB 2; Length 484;
Best Local Similarity 36.6%; Pred. No. 8.3e-48;
Matches 139; Conservative 77; Mismatches 117; Indels 47; Gaps 9;

| | | | |
|----|-----|--|-----|
| Qy | 33 | SPAGLGERVH-VLVLSWRSGSSFFVGOLFQHPDVFYLMPEAWHVDTLTSGSAPALHMA | 91 |
| Db | 110 | APEGVGDKRHWMYVFTTWRSQSPFFBELNQNPVEFLYEPVHWQKLYPGDAVSLQA | 169 |
| Qy | 92 | VRDLIRSVFLCDMDVFDAYLP---WRRNISDL--FQWAVSRALCSPPVCEAFARG--NIS | 144 |
| Db | 170 | ARDMLSALYRCDLISVFQLYSPAGSGGRNLTTLGIFGAATNKVCSPLCPAYRKEVVGLV | 229 |
| Qy | 145 | SEEVCKPLCATRPFGLAQACSSYSHVVLKEVRFNLOVLYPLLSDPALNRIVHLVRDP | 204 |
| Db | 230 | DDRVCCK-CPQRLARPEECRKYRTLVIKGVRFVDAVLAPLLRDPALDLKVIHLVRDP | 288 |
| Qy | 205 | RAVLRSREQTAKALARDNGIVLGTNGTWEADPR----- | 238 |
| Db | 289 | RAVASSIRSRHGLIRESLQVRSR-----DPAHRMPFLEAAGHKLGAKEGVGGPAD | 342 |
| Qy | 239 | ---LRVNVVECRSHVRIAEALHKPPFPFLQDRYRLVRYEDLARDPLTVIRELYAFTGLGL | 295 |
| Db | 343 | YHALGANEVICNSWAKTLQTAL-QPFDWLQGHLYLVRYEDLVGEFVKTLRRVDFVGLLV | 401 |
| Qy | 296 | TPQLQTVIWNITHSGGPGARREAFKTSRDALSVSQAWRHTLPFAKIRRVLOELCGALQL | 355 |
| Db | 402 | SPEMEQFALNMTSGSGSSK--PFWASARNATQANAWRTALTQQIKQVSEFCYQPMNV | 459 |
| Qy | 356 | LGYSRVSELEQORDLSLDLL 375 | |
| Db | 460 | LGYSRVNSPEEVKDLSTLL 479 | |

```
RESULT 2
JC7351
N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) [validated] - human
C:Species: Homo sapiens (man)
C:Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 09-Jul-2004
C:Accession: JC7351
R;Uchimura, K.; Fasakhany, F.; Kadamatsu, K.; Matsuoka, T.; Yamakawa, N.;
Biochem. Biophys. Res. Commun. 274, 291-296, 2000
A;Title: Diversity of N-acetylglucosamine-6-O-sulfotransferases: Molecular cloning of a
A;Reference number: JC7350; MUID:20374462; PMID:10913333
A;Accession: JC7351
A;Molecule type: mRNA
A;Residues: 1-486 <UCH>
A;Cross-references: UNIPROT:Q95667; DBJ:AB040711
C;Comment: This enzyme, having a type II transmembrane topology and N-linked glycosylation
sulfate linkage formation. This enzyme is involved in synthesis of L-selectin ligand in
C;Keywords: glycolysis; sulfate transport; sulfotransferase; transmembrane protein

Query Match 28.2%; Score 580; DB 2; Length 486;
Best Local Similarity 34.8%; Pred. No. 2.7e-42;
Matches 139; Conservative 56; Mismatches 137; Indels 68; Gaps 8;

QY 4 PRFSSTVMSLLMVQTIILVLSRVQVPSPPAGLGERVHVLVLSWSRSGSFVGLFSQH 63
DDB 83 PRFPSNL-----SGAVCEAVSR-----EKQHYVHATWRTGSSFLGELFNQH 124
QY 64 PDVFLMEPAHWVDLTLSQSSAPALHMAVRDLRSVFLCDMVFDAYLP----- 112
DDB 125 PDVFLYEPHMLWQALPGDABSLQCALRDLMSLFRCDVSLRLYAPPGDPAARAPDT 184
QY 113 WRNNSDLFOWAVSRALCSPVPC-----EAFAGNISSEEVCKPLCATRPFGLAQACSSY 168
DDB 185 ANLTALFWRNTKVICSPPLCPGAPRAAEVGLVEDTACERSCPPVAIRALEACRKY 244
QY 169 SHVVLKEVRFNQLVPLSDPALNRIHVLVRDPRAVLSRSEQTAKALARDNGIVL-- 226
DDB 245 PVVVKDRLDLGLVPLLRDPLGLNKVQLFRDPRAVHNSLRKSGRLRESIQVLR 304
QY 227 -----GTNGTWVEADPR-----LRVNVNEVCRSHVRIAEALH 258
DDB 305 RQRGRDRFHRVLLAHGVGAREPGOSRALPAAPRADFFLTGALEVICEAWLRDLFARGA-- 362
QY 259 KPPEFLQDRVRLVRYEDLARDPLTVIRELYAFTGLTLPQLOTWIHNTGSGPGARREA 318
DDB 363 --PAWLRRRLVRLVRYEDLVQPAQRLLRFLRGLRALAALDAFALNMTGGAAYGADR-P 419
QY 319 FKTTSRDALSVQAWRHRLPPAKIRRVQELCGGALQLLGY 358
DDB 420 FHLSDAREAVHAWRERLSREQVRQVEAACAPAMRLLAY 459

RESULT 3
JC7350
N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) [validated] - mouse
C:Species: Mus musculus (house mouse)
C:Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 09-Jul-2004
C:Accession: JC7350
R;Uchimura, K.; Fasakhany, F.; Kadamatsu, K.; Matsuoka, T.; Yamakawa, N.;
Biochem. Biophys. Res. Commun. 274, 291-296, 2000
A;Title: Diversity of N-acetylglucosamine-6-O-sulfotransferases: Molecular cloning of a
A;Reference number: JC7350; MUID:20374462; PMID:10913333
A;Accession: JC7350
A;Molecule type: mRNA
A;Residues: 1-484 <UCH>
A;Cross-references: UNIPROT:Q95N80; UNIPROT:Q95P78; DBJ:AB040710
C;Comment: This enzyme, having a type II transmembrane topology and N-linked glycosylation
sulfate linkage formation. This enzyme is involved in synthesis of L-selectin ligand in
C;Keywords: glycolysis; sulfate transport; sulfotransferase; transmembrane protein

Query Match 28.1%; Score 577; DB 2; Length 484;
Best Local Similarity 34.5%; Pred. No. 4.8e-42;
Matches 135; Conservative 57; Mismatches 141; Indels 58; Gaps 8;
```

```
QY 31 PSSPAGLG-----ERVHVLVLSWSRSGSFVGLFSQHDPDVFYLMPEPAHWVDLTLSQ 82
DDB 82 PPSGNLSAVGAVTQEKQHYVHATWRTGSSFLGELFNQHPDVFYLMPEHMLWQALYLP 141
QY 83 GSAPALHMAVRDLRSVFLCDMVFDAY-----LPMERNISD--LFQMAVSALCS 131
DDB 142 GDAESLQALRDLMSLPRCDVSLRLYAQPGDPCGERAPDSANLTTAMLFWRNTRKVC 201
QY 132 PIVCEAFARG-----NISSEVCKPLCATRPFGLAQACSSYSHVVLKEVRFNQLVLYLP 187
DDB 202 PPLCPAARARADVGLVEDKACESTCPPVSLRALEAECKRYPVVWIKOVRLLDLGLVLP 261
QY 188 LSDPALNLRIVHLVRDPRAVLSRSEQTAKALARDNGIVLGTN-----GTWVEAD 236
DDB 262 LRDPLGLNLKVQLFRDPRAVHNSLRKSGRLRESIQVLRTRQGRGHFHRVLLAHGVDA 321
QY 237 P-----LRVNVNEVCRSHVRIAEALHKKPPPPFQDRYRLVRYEDIA 277
DDB 322 PGQARALPSAPRADFFLTSALEVICEAWLRDLFTRGA----PAWLRRRLVRLVRYEDLV 377
QY 278 RDPLTVIRELYAFTGLTLPQLOTWIHNTGSGPGARREAFKTTSRDALSVQAWRHRL 337
DDB 378 WQFOAQLRLLRFLSGRLTLAALDAFANMTGSAAYGADR-PFHLSDAREAVHVMRRL 436
QY 338 PFAKIRRVQELCGGALQLLGYRSVHSELEOR 368
DDB 437 SOEQVRQVETACAPAMRLLAYPRSGDERDRK 467

RESULT 4
AS7397
chondroitin 6-sulfotransferase (EC 2.8.2.17) precursor - chicken
C:Species: Gallus gallus (chicken)
C:Date: 23-Feb-1996 #sequence_revision 23-Feb-1996 #text_change 09-Jul-2004
C:Accession: AS7397
R;Fukuta, M.; Uchimura, K.; Nakashima, K.; Kato, M.; Kimata, K.; Shinomura, T.; Habuchi,
J. Biol. Chem. 270, 18575-18580, 1995
A;Title: Molecular cloning and expression of chick chondrocyte chondroitin 6-sulfotransfe
A;Reference number: AS7397; MUID:95355490; PMID:7629189
A;Accession: AS7397
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-458 <FUK>
A;Cross-references: UNIPROT:Q92179; GB:D49915; NID:G971262; PIDN:BAA08655.1; PID:G971263
C;Superfamily: chondroitin 6-sulfotransferase
C;Keywords: sulfotransferase

Query Match 24.1%; Score 495; DB 2; Length 458;
Best Local Similarity 36.1%; Pred. No. 5.8e-35;
Matches 129; Conservative 59; Mismatches 125; Indels 44; Gaps 13;

QY 40 RVHVLVLSWSRSGSFVGLFSQHDPDVFYLMPEPAHWVDLTLS--OGSAPALHMAV--RDL 95
DDB 112 RRHVLMMATRTTSGSFVGEFFNQGNIFYLFEPLMHIERTVTFEPGAGANAVGSALVYRDV 171
QY 96 IRSVFLCDMVFDAYL---PWRNNSDLFQMAVSALCSPVPC-----EAFAGNISSE 147
DDB 172 LQQLLLCDLYLESFISPAPEHLLTAALFRGSSHSLCEEPVCTPSLKKVFEKHCNRR 231
QY 148 VCKPLCATRPFGLAQACSSYSHVVLKEVRFNQLVPLSDPALNLRIVHLVRDPRAV 207
DDB 232 -CGPLNIT-----LAAEACRKHQKMAKTVRIQLSEFLQPLAEDPRDLRLIQLVRDPRAV 286
QY 208 LRSREQTAKALARDNGIVLGTNGTW----VEADPLRLVVNEV-----CRSHVRIAEAL 257
DDB 287 LVSRMVAFS-----GKYESNKKWAAEEAPLQ-EDEVQLRNGNCESIRLSAELGL 335
QY 258 HKPPEFLQDRYRLVRYEDLARDPLTVIRELYAFTGLTLPQLOTWIHNTGSGPGARRE 317
DDB 336 -RQPRWLGRVYLVRYEDVARAPLRALEMYRFAGIHPFPQVEEVIIRANTQAPQDS---N 391
QY 318 AFKTTSRDALSVQAWRHRLPPAKIRRVQELCGGALQLLGYRSVHSELEOR 372
```

| Query Match | 6.1% | Score 126; | DB 2; | Length 388; |
|-----------------------|-------|-------------------------|-----------------------|-------------------------------------|
| Best Local Similarity | 23.1% | Pred. No. 0.0046; | | |
| Matches | 79; | Conservative | 50; | Mismatches 131; Indels 82; Gaps 17; |
| Qy | 43 | VLVLSSWRSGSSFFVQGLF---- | SOH--PDVFVLMPEAWHWDLT | SGSAPALHMAVRDLIR 97 |
| Db | 83 | IFVGHWRGTGTTLHLLVVD | DRHTGPTGYECLAPHFL---- | LTEWFAFYVE----- 131 |
| Qy | 98 | SVFLCD-----MDVDEYALP | WRRNISDLFQWAVSRALCP | VPVCEAFARGNISSEVCKPLC 153 |
| Db | 132 | --FLVSKHRAMND-- | -LSLHHPOQDEPFVCM-QGL | SPSYLTIAFPNRPQVEEY---- 181 |
| Qy | 154 | ATRPFGLAQBACSSYSHVV | LKEVFPFNLOVY----- | PLSDPALNLR----- 197 |
| Db | 182 | -----LDLEQVAPRELEI | WKRTLFRFVQVYFRERKTV | ILKNPTHFSFRIKVLEVPQA 235 |
| Qy | 198 | --VHLVDPDRAVLRSRQTA | KALARDNGIVLGTNGTWEAD | PLRVVNEVCRSHVRIAE 255 |
| Db | 236 | KFTHIVRDPVYVYPTSTH | LKALYRIHGL----- | QQPTFQGLD--KVSYVDLYRKLDG 290 |
| Qy | 256 | ALHKPPPPFLQDRYRLV | RYEDLARPPLTVIRELYAFT | GLG-----LTPQLQTWIHNITHSG 311 |
| Db | 291 | RELVDP-----TRFVEL | RYEDLIGPEGQLRLYQHLG | LDGDFCYLPRLQRYLAD--HAD- 343 |
| Qy | 312 | PGARREAFKTSRDALS | VSQAW-----RHTLPFAKIR | 343 |
| Db | 344 | --YKTSNGVOLTVEORAI | VEHGWBIIDRYGYDRHTPE | PARLR 383 |

DB ---1TSLTRHV-----WRBNYCACGNALRDCSFWSSVRKREWSGQDFGLMB 81
38 ---1TSLTRHV-----WRBNYCACGNALRDCSFWSSVRKREWSGQDFGLMB 81

T35999
probable aminoglycoside acetyltransferase - *Streptomyces coelicolor*
C/Species: *Streptomyces coelicolor*
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C/Accession: T35999
R/Seeger, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, July 1999
A/Reference number: Z21574
A/Accession: T35999
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-262 <SSE>
A/Cross-references: UNIPROT:Q9XAC9; EMBL:AL096839; PIDN:CAB50752.1; GSPDB:GN00070; SCOEDE
A/Experimental source: strain A3 (2)
C/Genetics:
A/Gene: SCOEDEB:SCC22.09
C/Summary: Family: *Escherichia coli* aminoglycoside N3'-acetyltransferase

| | | | |
|----|-----|--|-----|
| Qy | 225 | VLGTNGTWVEADPRLRVVNEVCRSHVRLAEALHKKPPSELQDRYRLRVRYEDLEARDPLTVI | 284 |
| | | : : : : : : | |
| Db | 57 | VLGPDGTLV-----VPTQTGDSLSDPALTSPVPVEEWETVRAATPAYDEL--- | 102 |
| | | : : : : : : | |
| Qy | 285 | RELYAFTGLGLTPO--LQTWIHNITHGSGPGARREAFKTTTSRDALSVSQAWRHLLPFAKIR | 343 |
| | | : : : : : | |

Db 103 --ITPSRGVGVVETRTW-----PGARRSAHPQTSFAALGG-----R 138

Qy 344 RVQELCGGALQL-LGYRSVHSELEQDLSLDLLPRGMD5 382

Db 139 AAEVVAGHATDCRIGERSPLAALERLDARV-LLLGAGYDA 177

RESULT 8

AG3174

acetolactate synthase [imported] - Agrobacterium tumefaciens (strain C58, Dupont) plasmid

C;Species: Agrobacterium tumefaciens

C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 16-Aug-2004

C;Accession: AG3174

R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.

erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutuyavin, T.; Levy, R.; Li, M.; McClell

i Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,

ster, E.W.

A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A;Reference number: AB2577; MUID:21608550; PMID:11743193

A;Accession: AG3174

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-571 <KUR>

A;Cross-references: UNIPROT:O8UKH9; GB:AE008687; PIDN:AAJ45813.1; PID:gl77433551; GSPDB:G

A;Experimental source: strain C58 (Dupont)

C;Genetics:

A;Gene: atrC

A;Genome: plasmid

C;Superfamily: Acetolactate synthase, large subunit/pyruvate oxidase; thiamin pyrophosph

Query Match 4.8%; Score 99; DB 2; Length 571;

Best Local Similarity 19.5%; Pred. No. 1.7;

Matches 70; Conservative 47; Mismatches 114; Indels 128; Gaps 14;

Qy 81 SOGSA-PALHMAVRDLRSVFLCDMDVFDAYLPWRRNISDLFOWAYSR----- 127

Db 151 SSGRAGFAVLLPADLRSV-----STTSRRRTNNLGHWPDLDRSPADKDLIDIA 201

Qy 128 -ALCSPVPCEAFARGNISSEVECKPLCATRPFGLAQEACS-----SYSHVLKVEVRFN 180

Db 202 KATANARAPVVIAGGIGHSG-----AAQRLQLQDECIPVFTTMMKGSDVEWHPLS 255

Qy 181 LOVLYPLSDPALNRLVHLVRDPAVLRSREQTAKALADNGIVLGT-----NGT--WVE 234

Db 256 GGVLASLVGPKSLGRHTVEIVREADLV-----VLIGTRTNQNGTDNRQ 299

Qy 235 ADPLRLRVNEVCRSHVRIAEAAHKPPFLQDRVRLVRYEDLARDPLTVIRELYAFTGLG 294

Db 300 YSPDAQIHH-----IDTDQEVGRNFPEARILVGDARETLAGURERIRLCDLH 346

Qy 295 LTPQLQPTWIHNITHGSGFGARREAFKTTSDAL--SVSQAWR-----HTLPFA 340

Db 347 L-----RTVSREAVCNRLRESWRKFNDDRRRGYSSSEASPLR 382

Qy 341 KIRRVQELG-----GALQLIGYRSVISELEQDLSLDLLPRGMD5FKW 385

Db 383 PERIMAELOGVIDENTTTVADASYSSMWVLGQLRISSERTR-----VLTFRGLAGLW 435

RESULT 9

F87316

conserved hypothetical protein CC0543 [imported] - Caulobacter crescentus

C;Species: Caulobacter crescentus

C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Jul-2004

C;Accession: F87316

R;Niernan, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.

B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon

n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A;Title: Complete Genome Sequence of Caulobacter crescentus.

A;Reference number: AB7249; MUID:21173698; PMID:11259647

A;Accession: F87316

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-856 <STO>

A;Cross-references: UNIPROT:Q9AAQ3; GB:AE005673; NID:gl3421732; PIDN:AAK22530.1; GSPDB:GF

C;Genetics:

A;Gene: CC0543

Query Match 4.8%; Score 98; DB 2; Length 856;

Best Local Similarity 19.5%; Pred. No. 3.5;

Matches 78; Conservative 53; Mismatches 152; Indels 118; Gaps 19;

Qy 39 ERVHVLVLSWSRSGS-----SFVQLFSQHPDVFYLMPEAWHV--W-----DTLS 81

Db 470 EMIH-----ANDKGARNVIANVGDIKPAEIGTSHFLEMAWDIDRWGKSQKQFLSDWTA 524

Qy 82 QGSAPALHMAVRDLRSVFLCDMDVFDAYLPWRRNISDLFOWAVSRALCSPPVCEAPARG 141

Db 525 RNLLGPALAAKATDLDRLRYRLNPFRRPEHLEW-----PPVAENRHLS 566

Qy 142 NISSEVECKPLCATRPFGLAQEACSSYSYSHVLKVEVRFNQLVLYPLSDPALNRLVHLV 201

Db 567 SYTPKEVSARLRAFRTLVAETKATSQQVPTLQDA--WFEL--VEFPIRISAAANMRFFAAE 624

Qy 202 RDPRAVLRSREQTAKALADNGIVLGTNGTWVEADPRLRVVNEVCRSHVRIAEAAHL--K 259

Db 625 R-YNALIDGRQ-----AMARSAG-----GAAVEAQAEITALTD--RFHQIAGGKWRWFM 671

Qy 260 PPPFLQDRVRLVRYED-----LARDPLTVIRELYAFTGLGLTLPOLQPTWIHNITHGSGP 312

Db 672 PEEPADSQWRIYRARPIPLPGAALTADPAFLAE-----VDGTLFAGSP 715

Qy 313 GARREAFKTTSDALSVSQAWRHTLPFAKIRR-----VQELCGGALQL----- 355

Db 716 VFEAEAFK-----ANRGWRFV---EGVGRGDGVMIAAAGADLTLEVEVKTEGRSUR 764

Qy 356 LGYRSVHSELEQDLSLDLL-----PRGMD5FKWA 386

Db 765 IGVLPFPFGQGGETALDVSINGEAPQVRSWPRAVGSPAWA 805

RESULT 10

A64878

peptide transport periplasmic protein sapA precursor - Escherichia coli (strain K-12)

C;Species: Escherichia coli

C;Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 09-Jul-2004

C;Accession: A64878

R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Coj

A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A;Title: The complete genome sequence of Escherichia coli K-12.

A;Reference number: A64720; MUID:97426617; PMID:9278503

A;Accession: A64878

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-547 <BLAT>

A;Cross-references: UNIPROT:Q47622; GB:AE000227; GB:U00096; NID:gl787543; PIDN:AAC74376.J

A;Experimental source: strain K-12, substrain MGL655

C;Genetics:

A;Gene: sapA

C;Superfamily: dipeptide transport protein

C;Keywords: periplasmic space; transport system

F;1-21/Domain: signal sequence #status predicted <SIG>

F;22-547/Product: peptide transport periplasmic protein sapA #status predicted <MAT>

Query Match 4.6%; Score 95; DB 2; Length 547;

Best Local Similarity 22.0%; Pred. No. 3.5;

Matches 109; Conservative 54; Mismatches 170; Indels 162; Gaps 27;

Qy 11 MSLMLMVQTGIL-----VFLVSRQVPS---SPAGLGRVHVVL 46

Db 4 VLSSLLIVAGLVSGOIAAPSPHADIIRDSDGFVYCVSGQVNTFNPSKASSGLIVDTL-- 61

Db 596 QSISKAGIVTSIQARCTVIAANPIGRDPSLTFSE-----NVDLTEPIIS-- 644
 Qy 192 ALNLRIVHLVRD-----PRAVLRSREQTAKALARDNGIVLGT-----NGTWVE 234
 Db 645 --RPDILCVVRDTPVQDEMLARFVGVSHRVHHPNSKEEGLANGSAAEPAMPNTYGV 702
 Qy 235 ADPLRVNVECRSHVRIAEAAHKPPFPLODRYELVR-YEDLARDPLTVIRELYAFTGL 293
 Db 703 PLPQ-----EVLKKYIIYAKERVH-PKLNQMDQDKVAKWYSDLRKESMAT----- 746
 Qy 294 GLTPOLQTVHNIITHGSGPGA-----RREAFKTTGRDALSVSQAWRHLLPFAKIRRV 345
 Db 747 GSPITVVRHIESMGGGPRAPHSAGLCDRRRRQGHGPRDA---GELHRT-----EVQRH 799
 Qy 346 QELCGGALQLGYRSVHSEL 365
 Db 800 RSMKRTFARYLSPRRDNNEL 819

 RESULT 12
 G90862
 hypothetical protein Ecs1871 [imported] - Escherichia coli (strain O157:H7, substrain R1N
 C:Species: Escherichia coli
 C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
 C:Accession: G90862
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
 Gasawara, N.; Yasunaka, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genom
 A:Reference number: A99629; MUID:21156231; PMID:11298796
 A:Accession: G90862
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-547 <HAY>
 A:Cross-references: UNIPROT:O8X7F3; GB:BA000007; PIDN:BAH35294.1; PID:g13361336; GSPDB:G
 A:Experimental source: strain O157:H7, substrain R1MD 0509952
 C:Genetics:
 A:Gene: Ecs1871
 C:Superfamily: dipeptide transport protein

 Query Match 4.5%; Score 93; DB 2; Length 547;
 Best Local Similarity 22.0%; Pred. No. 5.3;
 Matches 109; Conservative 53; Mismatches 171; Indels 162; Gaps 27;

 Qy 11 MSLMLVQTGIL-----VFLVSRQVPS---SPAGLGERVHVLVL 46
 Db 4 VLSELLVITAGLVSGQAIAPESPAPHADIRDSGFVYCVSQVNTFNPSKASSGLIVDTL-- 61
 Qy 47 SSWRSGSFVQQLFSQHDPDLYLMEPAKHVWDTLSQGSAPALHMAVRDL----- 95
 Db 62 ---AAQFYDLRLDLDVPTTYRLMPELAESWEVLNMGATYRFHLR-RDVPFQKTAFTPT 115
 Qy 96 ---IRSVFLCDMDVFDAYLDPWRNIS-----DLFQWAVSRALCSPPVCEAFARNIS 144
 Db 116 RKNADDVVTFFQRIFDNRNFW-HNVGNSNPPYFSLQFA-----DNVK 158
 Qy 145 SEEVCKPLCATRPFGLAQEAQSSYSHVVLKEVRFNQLVLYPLLSDFALNLRIVHLVRDP 204
 Db 159 S--VRKLDNHTVEFLAQPDASFVHLHATHYASVMSAEYARKLEKED---RQQLDRQP 212
 Qy 205 -----RAVLRSREQ-----TAKALARDNGIVLTNGT-----W 232
 Db 213 VGTGPQLSEYRAGQIFRLQRHDDFWRGKPLMPQVVDLGGGTGRSLKLLTGECDVLAW 272
 Qy 233 VEA-----DPLRVNVECRSHVRIAEAAHKPPFPFLOD---RYLVRVEDLARDPL 281
 Db 273 PAASQLSLTRDDPRLRL---TLRPGMNVAFLAFNTAKPPLNPNVPRHALA----LAINNQ 325
 Qy 282 TVIRELY---AFTGLGLTPQLQTVIHN-----ITHGSGPGARREAFKTTSRDALSV----- 329
 Db 326 RLMSIYYGTAEATAASILPR-ASWAYDNEAKITE-YNPAKSREQLKLLGLENTLKLWVP 383
 Qy 330 --SQAWRHHTLPFAKIRRVQELCGGALQLLGYSRVHSELEQR-----DLSLDILLPRGM 380

Db 384 TRSQWNPSP-PL-----KTAELIQDMAQGVKVVIVPVEGRFQEARLMDMSHDLTLS--- 435
 Qy 381 DSFKWASSTKQPES 395
 Db 436 ---GWATDS-NDPDS 446

RESULT 13
 B85756
 hypothetical protein sapA [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
 C;Species: Escherichia coli
 C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
 C;Accession: B85756
 R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A;Reference number: A85480; MUID:21074935; PMID:11206551
 A;Accession: B85756
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-547 <STO>
 A;Cross-references: UNIPROT:O8X7F3; GB:A805174; NID:G12515483; PIDN:AAG56510.1; GSPDB:G
 A;Experimental source: strain O157:H7, substrain EDL933
 C;Genetics:
 A;Gene: sapA
 C;Superfamily: dipeptide transport protein

Query Match 4.5%; Score 93; DB 2; Length 547;
 Best Local Similarity 22.0%; Pred. No. 5.3;
 Matches 109; Conservative 53; Mismatches 171; Indels 162; Gaps 27;
 Qy 11 MSLSLMVTQTGIL-----VFVLSRVQPS---SPAGLGERVHVVL 46
 Db 4 VLSSLLVIAGLVGQAIAAPESPPIADIRDSFGYVCVSGQVNTFNPSSKSSGLIVDTL-- 61
 Qy 47 SSWRSRGSFVGLFSQHPDVFYLMPEAWHVDTLSCGSAPALHMAVVDL----- 95
 Db 62 -----AAQFDRLLDVDPYRYLMPDLAESEVLDNGATYRFHLR-RDVPFQKTAMFTPT 115
 Qy 96 ----IRSVFLCDMDVFDAYLFWRRNIS-----DLFQWAVSRALCSPPVCEAFARGNIS 144
 Db 116 RKNMADDVFTFORIFDRNPNW-HNVNGSNFPYFDSLQFA-----DNVK 158
 Qy 145 SEEVCKPLCATRPGLAQACSSYSHVVLKEVRFNLOVLYPLSDPALNLRIVHLVRDP 204
 Db 159 S--VRKLDNHTVPEPLAQDPASFLWLHATHYASVMSAEYARKLEKED-----ROEQLDROP 212
 Qy 205 -----RAVLRREQ-----TAKALARDNGIVLGTNGT-----W 232
 Db 213 VGTGPYQLSEYRAGQFIRLQRHDDFWRGKPLMPQVVVDLSGGTGRLSKLLTGECVDVLA 272
 Qy 233 VEA-----DPLRVVNEVCRSHVRIABAALHKPPPLQD---RYRLVRYEDLARDPL 281
 Db 273 PAASQLSILRDDRRLD---TLRPGMVAVLAFTAKPPLNPAVRHALA---LAINNQ 325
 Qy 282 TVIRELY---AFTGLGTPTQLOTHIN-----ITHGSGPGARRAEAFKTTSDALSV----- 329
 Db 326 RLMQSIYVGAETAASILPR-ASWAYDNEAKITE-YNPAKSREQLKLLGLENTLKLWVP 383
 Qy 330 --SQAWRHTLPFAKIRRVQELCGGALLGVRSVHSELEOR-----DLSLDLLPRGM 380
 Db 384 TRSQWNPSP-PL-----KTAELIQDMAQGVKVVIVPVEGRFQEARLMDMSHDLTLS--- 435
 Qy 381 DSFKWASSTKQPES 395
 Db 436 ---GWATDS-NDPDS 446

RESULT 14
 T03412
 malate synthase (EC 4.1.3.2) - maize

C;Species: Zea mays (maize)
 C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 09-Jul-2004
 C;Accession: T03412
 R;Paek, N.C.
 submitted to the EMBL Data Library, September 1994
 A;Reference number: Z14939
 A;Accession: T03412
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 1-559 <PAE>
 A;Cross-references: UNIPROT:P49081; EMBL:L35914; NID:G532624; PIDN:AAB04118.1; PID:G53262;
 A;Experimental source: strain TX5855; scutellum
 C;Genetics:
 A;Note: MS
 C;Function: <ACG>
 A;Description: catalyzes the aldol condensation of glyoxylate with acetyl-CoA to form mal
 A;Pathway: glyoxylate cycle
 C;Superfamily: malate synthase
 C;Keywords: carbon-carbon lyase; coenzyme A; oxo-acid-lyase

Query Match 4.5%; Score 93; DB 2; Length 559;
 Best Local Similarity 22.5%; Pred. No. 5.4;
 Matches 56; Conservative 30; Mismatches 95; Indels 68; Gaps 12;
 Qy 159 GLAQEACSSYSHVYLKEVRFNQLVL-----YPLLSDPALNLRIVHLVRDPRVLRRE 212
 Db 312 GMAQHFMRSYSHLLIHTCHRRGVHANGMAAQIPIKDDAAANEAALELVKDKL----RE 367
 Qy 213 QTAKALARDNGIVLGTNGTWVEADPLRVVNEVCRSHV-----RIAEALHKPPPLQDR 267
 Db 368 VRA-----GHDGTAAHPGLIPAIRVFEHGLGRPNQIGDAAGHEGAS----- 411
 Qy 268 YRLRVYEDLARDPLTVIRELYAFTGLTPO-----LQTWIHNITHGSGPGARRAEAFKTT 322
 Db 412 ----VKEDLIQPP-----RGARTVDGLRLNRVGVQVLAALW-----AGSGSVPLYNLMEDA 460
 Qy 323 SRDALSVSQAW---RHTLPF-----AKIRRVQELCGGALLGVRSVHSELEORDLSDLL 375
 Db 461 ATAETISRVQWQLRHGAALDAGGVEVRATPELLA-----RVVEEMARVEAEV--- 509
 Qy 376 LPRGMDSPFK 384
 Db 510 ---GPDPRF 515

RESULT 15
 T50744
 spheroidene monooxygenase [imported] - Rhodobacter sphaeroides
 N;Alternate names: crtA protein
 C;Species: Rhodobacter sphaeroides
 C;Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 09-Jul-2004
 C;Accession: T50744
 R;Choudhary, M.; Kaplan, S.
 Nucleic Acids Res. 28, 862-867, 2000
 A;Title: DNA sequence analysis of the photosynthesis region of Rhodobacter sphaeroides 2.
 A;Reference number: Z25222; MUID:20115911; PMID:10648776
 A;Accession: T50744
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-327 <CHO>
 A;Cross-references: UNIPROT:Q9RFD1; EMBL:AF195122; PIDN:AAF24288.1
 A;Experimental source: strain 2.4.1
 C;Genetics:
 A;Gene: crtA
 C;Function:
 A;Pathway: carotenoid biosynthesis
 C;Superfamily: spheroidene monooxygenase

Query Match 4.5%; Score 92.5; DB 2; Length 327;
 Best Local Similarity 21.7%; Pred. No. 3;
 Matches 73; Conservative 29; Mismatches 147; Indels 87; Gaps 14;
 Qy 9 TVMLSLMVTQTGILVFLVSRQVPSAPGLGERVHVVLVLS-WRSRGSFVGLFSQHPDVF 67

Db 3 TVTLSIFRFNEFEKRLWVLGQMTANKLGM-----HYLPKAKFWKMGSGTGGFTPKEN-- 56
Qy 68 YLMEPAWHVNDTLSQGSAPALHMAVRDLIRSVFLCDMDVFDAYLPWRRNISDLFQMAVSR 127
Db 57 -----WHVWSILAVW--PDEETARREVAES-----PIYORWTMADES-----T 94
Qy 128 ALCSPVPVEAFARGNISSEEVCKPLCATRPFGLAQEA-----SSYSHVV-- 172
Db 95 VLLQPTSAWGKWDCKEPFEPV-KPASDVRPIAALTRATVVKFAERFWGARAASHMIGR 153
Qy 173 LKEVRF-----FNLOVLYPLLSDDPALNLRIVHLVRDPPAVLRSREQTAKALARONGI 224
Db 154 NKDVVFKIGVGEVPFVQOVTFSIWPDAAARW-----SSPRGAGGPHGEAIKAVRAENWF 208
Qy 225 -----VLGTNGTWEADPRLRVNVCRSH-----VRIAEALHK 259
Db 209 KEELYARFQILGTIGKWEKDP-----VGEALTARPSEAPKPPAPAVAQAPAAEAPKA 264
Qy 260 PPPFLQDRYRLRYEDLARDPLTVIRELYAFTGLGL 295
Db 265 PAPVAEKPALAVENPKPAEPPKPVVEAPKPPRSRGL 300

Search completed: June 23, 2005, 08:50:11
Job time : 11.9593 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2005, 08:49:14 ; Search time 49.935 Seconds
(without alignments)
3041.886 Million cell updates/sec

Title: US-10-697-828-7
Perfect score: 2057
Sequence: 1 MRLPRFSTWMLSLMVQTG.....LPRGMSFKWASSTKQPES 395

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1717557 seqs, 384547976 residues
Total number of hits satisfying chosen parameters: 1717557

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : Published Applications AA:*
- 1: /cgn2_6/ptodata/1/pubppa/US07_PUBCOMB.pep.*
 - 2: /cgn2_6/ptodata/1/pubppa/PCT_NEW_PUB.pep.*
 - 3: /cgn2_6/ptodata/1/pubppa/US06_NEW_PUB.pep.*
 - 4: /cgn2_6/ptodata/1/pubppa/US06_PUBCOMB.pep.*
 - 5: /cgn2_6/ptodata/1/pubppa/US07_NEW_PUB.pep.*
 - 6: /cgn2_6/ptodata/1/pubppa/PCTUS_PUBCOMB.pep.*
 - 7: /cgn2_6/ptodata/1/pubppa/US08_NEW_PUB.pep.*
 - 8: /cgn2_6/ptodata/1/pubppa/US08_PUBCOMB.pep.*
 - 9: /cgn2_6/ptodata/1/pubppa/US09A_PUBCOMB.pep.*
 - 10: /cgn2_6/ptodata/1/pubppa/US09B_PUBCOMB.pep.*
 - 11: /cgn2_6/ptodata/1/pubppa/US09C_PUBCOMB.pep.*
 - 12: /cgn2_6/ptodata/1/pubppa/US09D_PUBCOMB.pep.*
 - 13: /cgn2_6/ptodata/1/pubppa/US10A_PUBCOMB.pep.*
 - 14: /cgn2_6/ptodata/1/pubppa/US10B_PUBCOMB.pep.*
 - 15: /cgn2_6/ptodata/1/pubppa/US10C_PUBCOMB.pep.*
 - 16: /cgn2_6/ptodata/1/pubppa/US10D_PUBCOMB.pep.*
 - 17: /cgn2_6/ptodata/1/pubppa/US10E_PUBCOMB.pep.*
 - 18: /cgn2_6/ptodata/1/pubppa/US10F_PUBCOMB.pep.*
 - 19: /cgn2_6/ptodata/1/pubppa/US11A_PUBCOMB.pep.*
 - 20: /cgn2_6/ptodata/1/pubppa/US11B_PUBCOMB.pep.*
 - 21: /cgn2_6/ptodata/1/pubppa/US11C_PUBCOMB.pep.*
 - 22: /cgn2_6/ptodata/1/pubppa/US11D_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|---------------------|
| 1 | 2057 | 100.0 | 395 | 16 | US-10-697-828-7 |
| 2 | 2057 | 100.0 | 395 | 16 | US-10-841-707-8 |
| 3 | 2042.5 | 99.3 | 418 | 9 | US-09-927-602-5 |
| 4 | 1727 | 84.0 | 395 | 9 | US-09-927-602-3 |
| 5 | 1708 | 83.0 | 395 | 9 | US-09-927-602-2 |
| 6 | 1708 | 83.0 | 395 | 15 | US-10-258-080-5 |
| 7 | 1708 | 83.0 | 395 | 16 | US-10-648-593-159 |
| 8 | 1708 | 83.0 | 395 | 16 | US-10-697-828-13 |
| 9 | 1536.5 | 74.7 | 390 | 9 | US-09-927-602-4 |
| 10 | 1536.5 | 74.7 | 390 | 16 | US-10-697-828-8 |
| 11 | 1333 | 64.8 | 394 | 16 | US-10-408-765A-2211 |

| | | | | | |
|----|--------|------|------|----|----------------------|
| 12 | 1017.5 | 49.5 | 386 | 9 | US-09-816-825-2 |
| 13 | 1017.5 | 49.5 | 386 | 13 | US-10-007-262-1 |
| 14 | 1015.5 | 49.4 | 380 | 16 | US-10-841-707-6 |
| 15 | 986.5 | 48.0 | 386 | 14 | US-10-427-631-11 |
| 16 | 712 | 34.6 | 169 | 9 | US-09-927-602-6 |
| 17 | 677 | 32.9 | 169 | 9 | US-09-927-602-7 |
| 18 | 647.5 | 31.5 | 484 | 14 | US-10-212-933-4 |
| 19 | 647.5 | 31.5 | 531 | 9 | US-09-833-790-255 |
| 20 | 647.5 | 31.5 | 531 | 16 | US-10-755-889-284 |
| 21 | 640 | 31.1 | 483 | 14 | US-10-212-933-2 |
| 22 | 636.5 | 30.9 | 530 | 16 | US-10-723-860-1409 |
| 23 | 563.5 | 27.4 | 479 | 13 | US-10-087-192-126 |
| 24 | 550 | 26.7 | 481 | 13 | US-10-087-192-123 |
| 25 | 549 | 26.7 | 411 | 14 | US-10-021-660-128 |
| 26 | 549 | 26.7 | 411 | 15 | US-10-211-462-97 |
| 27 | 549 | 26.7 | 411 | 16 | US-10-408-765A-395 |
| 28 | 549 | 26.7 | 411 | 16 | US-10-723-860-1544 |
| 29 | 529.5 | 25.7 | 171 | 9 | US-09-927-602-8 |
| 30 | 377 | 18.3 | 169 | 9 | US-09-927-602-9 |
| 31 | 351.5 | 17.1 | 174 | 9 | US-09-927-602-11 |
| 32 | 338 | 16.4 | 179 | 9 | US-09-927-602-10 |
| 33 | 126 | 6.1 | 387 | 14 | US-10-126-279-21 |
| 34 | 126 | 6.1 | 387 | 14 | US-10-286-606-21 |
| 35 | 126 | 6.1 | 387 | 16 | US-10-891-383-21 |
| 36 | 118 | 5.7 | 596 | 16 | US-10-697-828-9 |
| 37 | 118 | 5.7 | 1212 | 17 | US-10-479-472A-2 |
| 38 | 118 | 5.7 | 1222 | 16 | US-10-697-828-15 |
| 39 | 118 | 5.7 | 1222 | 16 | US-10-475-446-4 |
| 40 | 117 | 5.7 | 1207 | 16 | US-10-697-828-17 |
| 41 | 117 | 5.7 | 1207 | 17 | US-10-479-472A-4 |
| 42 | 100.5 | 4.9 | 754 | 15 | US-10-425-563-18 |
| 43 | 100.5 | 4.9 | 754 | 18 | US-10-972-789A-18 |
| 44 | 97.5 | 4.7 | 1261 | 16 | US-10-437-963-200710 |
| 45 | 95.5 | 4.6 | 502 | 15 | US-10-369-493-19513 |

ALIGNMENTS

RESULT 1
US-10-697-828-7
; Sequence 7, Application US/10697828
; Publication No. US20040185546A1
; GENERAL INFORMATION:
; APPLICANT: Rosen, Steven
; APPLICANT: Lee, Jin Kyu
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: Novel Glycosyl Sulfotransferases GST-4alpha, GST-4beta, & GST-6
; FILE REFERENCE: UCAL-138DIV
; CURRENT APPLICATION NUMBER: US/10/697,828
; CURRENT FILING DATE: 2003-10-29
; PRIOR APPLICATION NUMBER: 09/593,828
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: 60/144,694
; PRIOR FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 395
; TYPE: PRT
; ORGANISM: mus musculus
US-10-697-828-7

| | | | | |
|-----------------------|-----------------|---|-----------|-------------|
| Query Match | 100.0% | Score 2057; | DB 16; | Length 395; |
| Best Local Similarity | 100.0% | Pred. No. 9.4e-196; | | |
| Matches 395; | Conservative 0; | Mismatches 0; | Indels 0; | Gaps 0; |
| Qy | 1 | MRLPRFSTWMLSLMVQTGILVFLVSRQVPSPPAGIGERVHVLVLSWSRSGSFVGLGF | 60 | |
| Db | 1 | MRLPRFSTWMLSLMVQTGILVFLVSRQVPSPPAGIGERVHVLVLSWSRSGSFVGLGF | 60 | |
| Qy | 61 | SQLPDPVFLMEPAWHVWDTTISQGSAPALHMAVRDLIRSVFLCDMDVFDAYLPWRNRISDL | 120 | |

```
Db 61 SQHPDVFYLMPEAWHVDLTSQGSAPALHMAVRDLIRSVFLCDMDVFDAYLPWRRNISDL 120
Qy 121 FQAVSRALCSPVCEAFARGNISSEVECKPLCATRPFGLAQACSSYSHVVLKEVRFFN 180
Db 121 FQAVSRALCSPVCEAFARGNISSEVECKPLCATRPFGLAQACSSYSHVVLKEVRFFN 180
Qy 181 LQVLYPLLSDPALNLRIVHLVRDPRAVLRSEQTAKALARDNGIIVLGTNGTWVEADPRLR 240
Db 181 LQVLYPLLSDPALNLRIVHLVRDPRAVLRSEQTAKALARDNGIIVLGTNGTWVEADPRLR 240
Qy 241 VVNEVCRSHVRIAEALHKPPPLQDRYLRVRYEDLARDPLTVIRELYAFTGLGTPQLQ 300
Db 241 VVNEVCRSHVRIAEALHKPPPLQDRYLRVRYEDLARDPLTVIRELYAFTGLGTPQLQ 300
Qy 301 TWIHNTHGSGPGARREAFKTTSRDALSVSQAWRHTLPFAKIRRVQELCGGALQLLGYRS 360
Db 301 TWIHNTHGSGPGARREAFKTTSRDALSVSQAWRHTLPFAKIRRVQELCGGALQLLGYRS 360
Qy 361 VHSLEQORDLSLDLLPRGMDSFKWASSTEKQPS 395
Db 361 VHSLEQORDLSLDLLPRGMDSFKWASSTEKQPS 395

RESULT 2
US-10-841-707-8
; Sequence 8, Application US/10841707
; Patent No. US20040202649A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Minoru
; APPLICANT: Yeh, Jiunn-Chern
; APPLICANT: Hiraoka, Nobuyoshi
; TITLE OF INVENTION: Identification of the Meca-79 Antigen
; TITLE OF INVENTION: and Related Methods of Treating L-Selectin-Mediated
; TITLE OF INVENTION: Conditions
; FILE REFERENCE: P-LJ 4149
; CURRENT APPLICATION NUMBER: US/10/841,707
; PRIOR FILING DATE: 2004-05-06
; PRIOR APPLICATION NUMBER: US/09/569,320A
; PRIOR FILING DATE: 2000-05-11
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Mus musculus
; US-10-841-707-8

Query Match 100.0%; Score 2057; DB 16; Length 395;
Best Local Similarity 100.0%; Pred. No. 9.4e-196;
Matches 395; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MRLPRFSSTVMSLLMVQGTGILVFLVSRQVPSPAGLGERVHVVLSSWRSGSSFVGQLF 60
Db 1 MRLPRFSSTVMSLLMVQGTGILVFLVSRQVPSPAGLGERVHVVLSSWRSGSSFVGQLF 60
Qy 61 SQHPDVFYLMPEAWHVDLTSQGSAPALHMAVRDLIRSVFLCDMDVFDAYLPWRRNISDL 120
Db 61 SQHPDVFYLMPEAWHVDLTSQGSAPALHMAVRDLIRSVFLCDMDVFDAYLPWRRNISDL 120
Qy 121 FQAVSRALCSPVCEAFARGNISSEVECKPLCATRPFGLAQACSSYSHVVLKEVRFFN 180
Db 121 FQAVSRALCSPVCEAFARGNISSEVECKPLCATRPFGLAQACSSYSHVVLKEVRFFN 180
Qy 181 LQVLYPLLSDPALNLRIVHLVRDPRAVLRSEQTAKALARDNGIIVLGTNGTWVEADPRLR 240
Db 181 LQVLYPLLSDPALNLRIVHLVRDPRAVLRSEQTAKALARDNGIIVLGTNGTWVEADPRLR 240
Qy 241 VVNEVCRSHVRIAEALHKPPPLQDRYLRVRYEDLARDPLTVIRELYAFTGLGTPQLQ 300
Db 241 VVNEVCRSHVRIAEALHKPPPLQDRYLRVRYEDLARDPLTVIRELYAFTGLGTPQLQ 300
Qy 301 TWIHNTHGSGPGARREAFKTTSRDALSVSQAWRHTLPFAKIRRVQELCGGALQLLGYRS 360
Db 301 TWIHNTHGSGPGARREAFKTTSRDALSVSQAWRHTLPFAKIRRVQELCGGALQLLGYRS 360
```

```
Db 301 TWIHNTHGSGPGARREAFKTTSRDALSVSQAWRHTLPFAKIRRVQELCGGALQLLGYRS 360
Qy 361 VHSLEQORDLSLDLLPRGMDSFKWASSTEKQPS 395
Db 361 VHSLEQORDLSLDLLPRGMDSFKWASSTEKQPS 395

RESULT 3
US-09-927-602-5
; Sequence 5, Application US/09927602
; Patent No. US20020061562A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; APPLICANT: Akama, Tomoya O.
; TITLE OF INVENTION: Methods of Treating Macular Corneal
; TITLE OF INVENTION: Dystrophy
; FILE REFERENCE: P-LJ 4852
; CURRENT APPLICATION NUMBER: US/09/927,602
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 09/638,211
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 418
; TYPE: PRT
; ORGANISM: Mus musculus
; US-09-927-602-5

Query Match 99.3%; Score 2042.5; DB 9; Length 418;
Best Local Similarity 99.7%; Pred. No. 2.8e-194;
Matches 394; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 MRLPRFSSTVMSLLMVQGTGILVFLVSRQVPSPAGLGERVHVVLSSWRSGSSFVGQLF 60
Db 25 MRLPRFSSTVMSLLMVQGTGILVFLVSRQVPSPAGLGERVHVVLSSWRSGSSFVGQLF 84
Qy 61 SQHPDVFYLMPEAWHVDLTSQGSAPALHMAVRDLIRSVFLCDMDVFDAYLPWRRNISDL 120
Db 85 SQHPDVFYLMPEAWHVDLTSQGSAPALHMAVRDLIRSVFLCDMDVFDAYLPWRRNISDL 144
Qy 121 FQAVSRALCSPVCEAFARGNISSEVECKPLCATRPFGLAQACSSYSHVVLKEVRFFN 180
Db 145 FQAVSRALCSPVCEAFARGNISSEVECKPLCATRPFGLAQACSSYSHVVLKEVRFFN 204
Qy 181 LQVLYPLLSDPALNLRIVHLVRDPRAVLRSEQTAKALARDNGIIVLGTNGTWVEADPRLR 240
Db 205 LQVLYPLLSDPALNLRIVHLVRDPRAVLRSEQTAKALARDNGIIVLGTNGTWVEADPRLR 264
Qy 241 VVNEVCRSHVRIAEALHKPPPLQDRYLRVRYEDLARDPLTVIRELYAFTGLGTPQLQ 300
Db 265 VVNEVCRSHVRIAE-ALHKPPPLQDRYLRVRYEDLARDPLTVIRELYAFTGLGTPQLQ 323
Qy 301 TWIHNTHGSGPGARREAFKTTSRDALSVSQAWRHTLPFAKIRRVQELCGGALQLLGYRS 360
Db 324 TWIHNTHGSGPGARREAFKTTSRDALSVSQAWRHTLPFAKIRRVQELCGGALQLLGYRS 383
Qy 361 VHSLEQORDLSLDLLPRGMDSFKWASSTEKQPS 395
Db 384 VHSLEQORDLSLDLLPRGMDSFKWASSTEKQPS 418

RESULT 4
US-09-927-602-3
; Sequence 3, Application US/09927602
; Patent No. US20020061562A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; APPLICANT: Akama, Tomoya O.
; TITLE OF INVENTION: Methods of Treating Macular Corneal
; TITLE OF INVENTION: Dystrophy
; FILE REFERENCE: P-LJ 4852
; CURRENT APPLICATION NUMBER: US/09/927,602
```

;; CURRENT FILING DATE: 2001-08-09
;; PRIOR APPLICATION NUMBER: US 09/638,211
;; PRIOR FILING DATE: 2000-08-11
;; NUMBER OF SEQ ID NOS: 38
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 3
;; LENGTH: 395
;; TYPE: PRT
;; ORGANISM: Artificial Sequence
;; OTHER INFORMATION: synthetic construct
;; NAME/KEY: VARIANT
;; LOCATION: (1)...(395)
;; OTHER INFORMATION: Xaa = any amino acid
US-09-927-602-3

Query Match 84.0%; Score 1727; DB 9; Length 395;
Best Local Similarity 84.7%; Pred. No. 6.8e-163;
Matches 333; Conservative 16; Mismatches 44; Indels 0; Gaps 0;
Qy 1 MRLPRFSTVMSLLMVQGTGLVFLVSRQVPSPAGLGERVHVLVLSWSRSGSFVQQLF 60
Db 1 MRLPRFSTVXTXLLLAQTLLFLVSRPGSPAGGEXRVHVLVLSWSRSGSFVQQLF 60
Qy 61 SQHPDVFLMEPAWHVMDTILSQGSAPALHMAVRDLIRSVFLCDMDVDFDAYLPWRRNISDL 120
Db 61 SQHPDVFLMEPAWHVMDTILSQGSAAATLHMAVRDLXRSVFLCDMDVDFDAYLPWRRNISDL 120
Qy 121 FQAVSRALCSPPVCEAFARNISSSEVCKPLCATRPFGLAQACSSYSHVVLKEVRFN 180
Db 121 FQAVSRALCSPPVACSAFPRGXISSEVCKPLCARQPFGLAREACRSYSHVVLKEVRFN 180
Qy 181 LOVLYPLSDPALNLRIVHLVRDPRVLRSEQTAKALARDNGIVLGTNGTWVEADPRLR 240
Db 181 LOVLYPLSDPALNLRIVHLVRDPRVLRSEQTAKALARDNGIVLGTNGTWVEADPRLR 240
Qy 241 VNEVCRSHVRIAEALHKPPPLQDRYRLVRYEDLARDPLTVIRELYAFTGLGLTQQLQ 300
Db 241 VREVCSSHVRIAEALHKPPPLGRYRLVRFEDLAREPLAEIRALYAFGLTGLTQLE 300
Qy 301 TWIHNITHGSGPGARREAFKTTSDALSVSQAWRHTLPFAKIRRVQELCGALQLLGYRS 360
Db 301 AWIHNITHGSGPGARREAFKTSRNALNVSQAWRHLPFAKIRRVQELCGALQLLGYRP 360
Qy 361 VHSLEQRDLSDLLPRGMDSFKWASSTKQP 393
Db 361 VYSEDEQRDLSDLLPRGXDFXWASSTXXXXP 393

RESULT 5
US-09-927-602-2
; Sequence 2, Application US/09927602
; Patent No. US20020061562A1
; GENERAL INFORMATION:
; APPLICANT: Akama, Michiko N.
; TITLE OF INVENTION: Methods of Treating Macular Corneal
; FILE OF INVENTION: Dystrophy
; FILE REFERENCE: P-LJ 4852
; CURRENT APPLICATION NUMBER: US/09/927,602
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 09/638,211
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-927-602-2

Query Match 83.0%; Score 1708; DB 9; Length 395;
Best Local Similarity 83.0%; Pred. No. 5.3e-161;

Matches 328; Conservative 24; Mismatches 43; Indels 0; Gaps 0;
Qy 1 MRLPRFSTVMSLLMVQGTGLVFLVSRQVPSPAGLGERVHVLVLSWSRSGSFVQQLF 60
Db 1 MRLPRVSTAVTALLAQITLLFLVSRPGSPAGGEXRVHVLVLSWSRSGSFVQQLF 60
Qy 61 SQHPDVFLMEPAWHVMDTILSQGSAPALHMAVRDLIRSVFLCDMDVDFDAYLPWRRNISDL 120
Db 61 NQHPDVFLMEPAWHVMDTILSQGSAAATLHMAVRDLVRSVFLCDMDVDFDAYLPWRRNISDL 120
Qy 121 FQAVSRALCSPPVCEAFARNISSSEVCKPLCATRPFGLAQACSSYSHVVLKEVRFN 180
Db 121 FQAVSRALCSPPVACSAFPRGALSSSEAVCKPLCARQSFLLAREACRSYSHVVLKEVRFN 180
Qy 181 LOVLYPLSDPALNLRIVHLVRDPRVLRSEQTAKALARDNGIVLGTNGTWVEADPRLR 240
Db 181 LOVLYPLSDPALNLRIVHLVRDPRVLRSEQTAKALARDNGIVLGTNGTWVEADPRLR 240
Qy 241 VNEVCRSHVRIAEALHKPPPLQDRYRLVRYEDLARDPLTVIRELYAFTGLGLTQQLQ 300
Db 241 VREVCSSHVRIAEALHKPPPLGRYRLVRFEDLAREPLAEIRALYAFGLTGLTQLE 300
Qy 301 TWIHNITHGSGPGARREAFKTTSDALSVSQAWRHTLPFAKIRRVQELCGALQLLGYRS 360
Db 301 AWIHNITHGSGPGARREAFKTSRNALNVSQAWRHLPFAKIRRVQELCGALQLLGYRP 360
Qy 361 VHSLEQRDLSDLLPRGMDSFKWASSTKQPES 395
Db 361 VYSEDEQRDLSDLLPRGLNGTWTWASSTASHPRN 395
RESULT 6
US-10-258-080-5
; Sequence 5, Application US/10258080
; Publication No. US20040029125A1
; GENERAL INFORMATION:
; APPLICANT: Incyte Genomics, Inc.
; APPLICANT: POLICKY, Jennifer L.
; APPLICANT: HAPALIA, April J.A.
; APPLICANT: BURFORD, Neil
; APPLICANT: RING, Huijun Z.
; APPLICANT: LAL, Preeti G.
; APPLICANT: TRIBOULEY, Catherine M.
; APPLICANT: YAO, Monique G.
; APPLICANT: YUB, Henry
; APPLICANT: TANG, Y. Tom
; APPLICANT: ARVIZU, Chandra S.
; APPLICANT: DAS, Debopriya
; APPLICANT: SANJANWALA, Madhusudan M.
; APPLICANT: GANDHI, Ameet R.
; APPLICANT: REDDY, Roopa M.
; APPLICANT: KHAN, Farrah A.
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: RAMKUMAR, Jayalaxmi
; APPLICANT: GRIFFIN, Jennifer A.
; APPLICANT: AU-YOUNG, Janice K.
; TITLE OF INVENTION: DRUG METABOLIZING ENZYMES
; FILE REFERENCE: PI-0070 USN
; CURRENT APPLICATION NUMBER: US/10/258,080
; PRIOR FILING DATE: 2002-10-15
; PRIOR APPLICATION NUMBER: US 60/203,509
; PRIOR FILING DATE: 2000-05-11
; PRIOR APPLICATION NUMBER: US 60/202,234
; PRIOR FILING DATE: 2000-05-05
; PRIOR APPLICATION NUMBER: US 60/200,185
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/198,403
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: PCT/US01/11869
; PRIOR FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: US 60/197,590
; PRIOR FILING DATE: 2000-04-13
; NUMBER OF SEQ ID NOS: 20

```
; SOFTWARE: PERL Program
; SEQ ID NO 5
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20040029125A1 7472777CD1
US-10-258-080-5

Query Match      83.0%; Score 1708; DB 15; Length 395;
Best Local Similarity 83.0%; Pred. No. 5.3e-161;
Matches 328; Conservative 24; Mismatches 43; Indels 0; Gaps 0;

Qy 1 MRLPRFSTVMSLLMVQTGILVFLVSRQVPSPAGLGERVHVVLSSWRSGSFVGQLF 60
Db 1 MWLPRVSTAVTALLAQTFLLFLVSRPGPSPAGGEARVHVVLSSWRSGSFVGQLF 60
Qy 61 SQHPDVFLMEPAWHVMDTLTSGSAPALHMAVRDLIRSVFLCDMDVDFDAYLPWRRNISDL 120
Db 61 NQHPDVFLMEPAWHVMDTLTSGSAPALHMAVRDLIRSVFLCDMDVDFDAYLPWRRNISDL 120
Qy 121 FQAVSRALCSPPVCEAFARGNISSEEVCKPCLCATRPFGLAQEACSSYSHVVLKEVRF 180
Db 121 FQAVSRALCSPPVCEAFARGNISSEEVCKPCLCATRPFGLAQEACSSYSHVVLKEVRF 180
Qy 181 LOVLYPLLSDPALNLRIVHLVRDPRAVLRREQTAKALARDNGIVLGTNGTWVEADPRLR 240
Db 181 LOVLYPLLSDPALNLRIVHLVRDPRAVLRREQTAKALARDNGIVLGTNGTWVEADPRLR 240
Qy 241 VVNEVCRSHVRIAEAAHKPPPLQDRYRLVRYEDLARDPLTVIRELYAFTGLGLTPQLQ 300
Db 241 VVNEVCRSHVRIAEAAHKPPPLQDRYRLVRYEDLARDPLTVIRELYAFTGLGLTPQLQ 300
Qy 301 TWIHNITHGSGPGARREAFKTTSDALSVSQAWRHHTLPFAKIRRVQELCGGALQLLYGRS 360
Db 301 TWIHNITHGSGPGARREAFKTTSDALSVSQAWRHHTLPFAKIRRVQELCGGALQLLYGRS 360
Qy 361 VVSELEQRDLSDLLPRGMDSKWASSTEKOPES 395
Db 361 VVSEDEQRNLALDLVLRGLNGFTWASSTASHPRN 395

RESULT 7
US-10-648-593-159
; Sequence 159, Application US/10648593
; Publication No. US20040106132A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: IDENTIFICATION OF GENES FOR PREDICTING ACTIVITY OF COMPOUNDS THAT
; TITLE OF INVENTION: INTERACT WITH AND/OR MODULATE PROTEIN TYROSINE KINASES AND/OR
; FILE REFERENCE: D0273 NP
; CURRENT FILING DATE: 2003-08-26
; PRIOR APPLICATION NUMBER: US/10/648,593
; PRIOR FILING DATE: 2002-08-27
; NUMBER OF SEQ ID NOS: 557
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 159
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-648-593-159

Query Match      83.0%; Score 1708; DB 16; Length 395;
Best Local Similarity 83.0%; Pred. No. 5.3e-161;
Matches 328; Conservative 24; Mismatches 43; Indels 0; Gaps 0;

Qy 1 MRLPRFSTVMSLLMVQTGILVFLVSRQVPSPAGLGERVHVVLSSWRSGSFVGQLF 60
Db 1 MWLPRVSTAVTALLAQTFLLFLVSRPGPSPAGGEARVHVVLSSWRSGSFVGQLF 60
Qy 61 SQHPDVFLMEPAWHVMDTLTSGSAPALHMAVRDLIRSVFLCDMDVDFDAYLPWRRNISDL 120
Db 61 NQHPDVFLMEPAWHVMDTLTSGSAPALHMAVRDLIRSVFLCDMDVDFDAYLPWRRNISDL 120
Qy 121 FQAVSRALCSPPVCEAFARGNISSEEVCKPCLCATRPFGLAQEACSSYSHVVLKEVRF 180
Db 121 FQAVSRALCSPPVCEAFARGNISSEEVCKPCLCATRPFGLAQEACSSYSHVVLKEVRF 180
Qy 181 LOVLYPLLSDPALNLRIVHLVRDPRAVLRREQTAKALARDNGIVLGTNGTWVEADPRLR 240
Db 181 LOVLYPLLSDPALNLRIVHLVRDPRAVLRREQTAKALARDNGIVLGTNGTWVEADPRLR 240
Qy 241 VVNEVCRSHVRIAEAAHKPPPLQDRYRLVRYEDLARDPLTVIRELYAFTGLGLTPQLQ 300
Db 241 VVNEVCRSHVRIAEAAHKPPPLQDRYRLVRYEDLARDPLTVIRELYAFTGLGLTPQLQ 300
Qy 301 TWIHNITHGSGPGARREAFKTTSDALSVSQAWRHHTLPFAKIRRVQELCGGALQLLYGRS 360
Db 301 TWIHNITHGSGPGARREAFKTTSDALSVSQAWRHHTLPFAKIRRVQELCGGALQLLYGRS 360
Qy 361 VVSELEQRDLSDLLPRGMDSKWASSTEKOPES 395
Db 361 VVSEDEQRNLALDLVLRGLNGFTWASSTASHPRN 395

RESULT 8
US-10-697-828-13
; Sequence 13, Application US/10697828
; Publication No. US20040185546A1
; GENERAL INFORMATION:
; APPLICANT: Rosen, Steven
; APPLICANT: Lee, Jin Kyu
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: Novel Glycosyl Sulfotransferases GST-4alpha, GST-4beta, & GST-6
; FILE REFERENCE: UCAL-138DIV
; CURRENT FILING DATE: 2003-10-29
; PRIOR APPLICATION NUMBER: US/10/697,828
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: 60/144,694
; PRIOR FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 395
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-697-828-13

Query Match      83.0%; Score 1708; DB 16; Length 395;
Best Local Similarity 83.0%; Pred. No. 5.3e-161;
Matches 328; Conservative 24; Mismatches 43; Indels 0; Gaps 0;

Qy 1 MRLPRFSTVMSLLMVQTGILVFLVSRQVPSPAGLGERVHVVLSSWRSGSFVGQLF 60
Db 1 MWLPRVSTAVTALLAQTFLLFLVSRPGPSPAGGEARVHVVLSSWRSGSFVGQLF 60
Qy 61 SQHPDVFLMEPAWHVMDTLTSGSAPALHMAVRDLIRSVFLCDMDVDFDAYLPWRRNISDL 120
Db 61 NQHPDVFLMEPAWHVMDTLTSGSAPALHMAVRDLIRSVFLCDMDVDFDAYLPWRRNISDL 120
Qy 121 FQAVSRALCSPPVCEAFARGNISSEEVCKPCLCATRPFGLAQEACSSYSHVVLKEVRF 180
Db 121 FQAVSRALCSPPVCEAFARGNISSEEVCKPCLCATRPFGLAQEACSSYSHVVLKEVRF 180
Qy 181 LOVLYPLLSDPALNLRIVHLVRDPRAVLRREQTAKALARDNGIVLGTNGTWVEADPRLR 240
Db 181 LOVLYPLLSDPALNLRIVHLVRDPRAVLRREQTAKALARDNGIVLGTNGTWVEADPRLR 240
Qy 241 VVNEVCRSHVRIAEAAHKPPPLQDRYRLVRYEDLARDPLTVIRELYAFTGLGLTPQLQ 300
Db 241 VVNEVCRSHVRIAEAAHKPPPLQDRYRLVRYEDLARDPLTVIRELYAFTGLGLTPQLQ 300
```

Qy 301 TWIHNTHGSGPGARRBAFTKTSRDALSVSQAWRHLLPFAKIRVQBLCGALQLLGYR 360
Db 301 AWIHNTHGSGPGARRBAFTKTSRNALNVSQAWRHLPFAKIRRVQBLCGALQLLGYR 360
Qy 361 VHSLEQRDLSDLLPRGMDSPKASSTKQPS 395
Db 361 VYSEQRNLALDLVLRGLNGFTWASSTASHPRN 395

RESULT 9

US-09-927-602-4
; Sequence 4, Application US/09927602
; Patent No. US20020061562A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; APPLICANT: Akama, Tomoya O.
; TITLE OF INVENTION: Methods of Treating Macular Corneal
; Dystrophy
; FILE REFERENCE: P-LJ 4852
; CURRENT APPLICATION NUMBER: US/09/927,602
; CURRENT FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 09/638,211
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 390
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-927-602-4

Query Match 74.7%; Score 1536.5; DB 9; Length 390;
Best Local Similarity 76.0%; Pred. No. 6.3e-144;
Matches 295; Conservative 30; Mismatches 62; Indels 1; Gaps 1;

Qy 1 MRLPRFSTVMSLLMVQTG-ILVFLVSRQVPSPPAGLGRVHVLVSSWRSGSFGQL 59
Db 1 MWLPFSKTVTVLLLAQTTCLLLFIISRPSPAGGEDRVHVLVSSWRSGSFGQL 60
Qy 60 FSQHPDVFLMEPAWHVWDTLSQSAPALHMAVRDLRSVFLCDMDVFDALPWRNITSD 119
Db 61 FSQHPDVFLMEPAWHVWDTLSQSSAATLHMAVRDLMSIFLCDMDVFDAYMPSRNL 120
Qy 120 LFQWAVSRALCSPVCEAFARGNISSEVCKPLCATRPFGLAQACSSYSHVVLKEVRFF 179
Db 121 FFWATSRALCSPACSAFPGTISKQDVCKTLCTROPFSLARACRSYSHVVLKEVRFF 180
Qy 180 NLQVLYPLSDPALNLRIVHLVRDPAVLRSREQTAKALARDNGIVLGTNGTWVEADPRL 239
Db 181 NLQVLYPLSDPALNLRIVHLVRDPAVLRSREAAAGPILARDNGIVLGTNGKWEADPHL 240
Qy 240 RVNEVCRSHVRIAEAAHLKPPFPLODRYRLVRYEDLARDPLTVIRELYAFTGLTPQL 299
Db 241 RLIREVCRSHVRIAEAAHLKPPFPPLGRYRLVRFEDLAREPLAEIRALYAFGLTLPQL 300
Qy 300 OTWNIHNTGSGPGARRBAFTKTSRDALSVSQAWRHLLPFAKIRVQBLCGALQLLGYR 359
Db 301 EAWIHNTHGSGIGKPIEAFHTSSRNARVNSQAWRHLPFTKILRVQEVCGALQLLGYR 360
Qy 360 SVHSEQRDLSDLLPRGMDSPKAS 387
Db 361 PVYSADQQRDLTLVLPRGPDHFSWAS 388

RESULT 10

US-10-697-828-8
; Sequence 8, Application US/10697828
; Publication No. US20040185546A1
; GENERAL INFORMATION:
; APPLICANT: Rosen, Steven
; APPLICANT: Lee, Jin Kyu
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: Novel Glycosyl Sulfotransferases GST-4alpha, GST-4beta, & GST-6

; FILE REFERENCE: UCAL-138DIV
; CURRENT APPLICATION NUMBER: US/10/697,828
; CURRENT FILING DATE: 2003-10-29
; PRIOR APPLICATION NUMBER: 09/593,828
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: 60/144,694
; PRIOR FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 390
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-697-828-8

Query Match 74.7%; Score 1536.5; DB 16; Length 390;
Best Local Similarity 76.0%; Pred. No. 6.3e-144;
Matches 295; Conservative 30; Mismatches 62; Indels 1; Gaps 1;

Qy 1 MRLPRFSTVMSLLMVQTG-ILVFLVSRQVPSPPAGLGRVHVLVSSWRSGSFGQL 59
Db 1 MWLPFSKTVTVLLLAQTTCLLLFIISRPSPAGGEDRVHVLVSSWRSGSFGQL 60
Qy 60 FSQHPDVFLMEPAWHVWDTLSQSAPALHMAVRDLRSVFLCDMDVFDALPWRNITSD 119
Db 61 FSQHPDVFLMEPAWHVWDTLSQSSAATLHMAVRDLMSIFLCDMDVFDAYMPSRNL 120
Qy 120 LFQWAVSRALCSPVCEAFARGNISSEVCKPLCATRPFGLAQACSSYSHVVLKEVRFF 179
Db 121 FFWATSRALCSPACSAFPGTISKQDVCKTLCTROPFSLARACRSYSHVVLKEVRFF 180
Qy 180 NLQVLYPLSDPALNLRIVHLVRDPAVLRSREQTAKALARDNGIVLGTNGTWVEADPRL 239
Db 181 NLQVLYPLSDPALNLRIVHLVRDPAVLRSREAAAGPILARDNGIVLGTNGKWEADPHL 240
Qy 240 RVNEVCRSHVRIAEAAHLKPPFPLODRYRLVRYEDLARDPLTVIRELYAFTGLTPQL 299
Db 241 RLIREVCRSHVRIAEAAHLKPPFPPLGRYRLVRFEDLAREPLAEIRALYAFGLTLPQL 300
Qy 300 OTWNIHNTGSGPGARRBAFTKTSRDALSVSQAWRHLLPFAKIRVQBLCGALQLLGYR 359
Db 301 EAWIHNTHGSGIGKPIEAFHTSSRNARVNSQAWRHLPFTKILRVQEVCGALQLLGYR 360
Qy 360 SVHSEQRDLSDLLPRGMDSPKAS 387
Db 361 PVYSADQQRDLTLVLPRGPDHFSWAS 388

RESULT 11

US-10-408-765A-2211
; Sequence 2211, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 66088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2211
; LENGTH: 394
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-2211

```
Query Match      64.8%; Score 1333; DB 16; Length 394;
Best Local Similarity 78.1%; Pred. No. 1.2e-123;
Matches 257; Conservative 19; Mismatches 47; Indels 6; Gaps 1;

Qy 1 MRLPRFSTVMSLLMVQTGILVLSRQVPSPAGLGERVHVLVLSWRSGSFVGQLF 60
Db 1 MMLPRVSTAVTALLAQTELLFLVSRPGPSPAGGEARVHVLVLSWRSGSFVGQLF 60

Qy 61 SQHPDVFLMEPAWHVMDTILSQSAPALHMAVRDLIRSVFLCDMDVFDAYLPWRRNISDL 120
Db 61 NOHPDVFLMEPAWHVMDTILSQSAAALHMAVRDLIRSVFLCDMDVFDAYLPWRRNISDL 120

Qy 121 FQAVSRALCSPPVCEAFARGNISSEBCKPLCATRPFGLAQACSSYSHVLKEVRFN 180
Db 121 FQAVSRALCSPPVCEAFARGNISSEBCKPLCATRPFGLAQACSSYSHVLKEVRFN 180

Qy 181 LOVLPLSDPALNLRIVHLVRDPAVLRVREQTAKALARDNGIVLGTNGTWVEADPRLR 240
Db 181 LOVLPLSDPALNLRIVHLVRDPAVLRVREQTAKALARDNGIVLGTNGTWVEADPRLR 240

Qy 241 VNEVCRSHVRIAEALHKPPFLQDRYLRVRYEDLARDPLTVIRELYAFTGLTLPQLQ 300
Db 241 LIREVCRSHVRIAEALHKPPFLQDRYLRVRYEDLAREPLAEIRALYFTGLTLPQLE 300

Qy 301 TWIHNTHGSG-----PGARREAFKTTTS 323
Db 301 AWIHNTHGSGASQSPSILRLGMRATS 329

RESULT 12
US-09-816-825-2
; Sequence 2, Application US/09816825
; Patent No. US20010051370A1
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107COSN
; CURRENT APPLICATION NUMBER: US/09/816,825
; CURRENT FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 09/045,284
; PRIOR FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-816-825-2

Query Match      49.5%; Score 1017.5; DB 9; Length 386;
Best Local Similarity 56.4%; Pred. No. 3e-92;
Matches 215; Conservative 43; Mismatches 112; Indels 11; Gaps 7;

Qy 1 MRLPRFSTVMSLLMVQTGILVF---LVSRQVPS-SPAGLGERVHVLVLSWRSGSFV 56
Db 1 MLLPK--KMKLLFLVSOQMAILALFFHMYSHNLSLWKQAQPERMHVVLVLSWRSGSFV 58

Qy 57 GOLFSQHPDVFLMEPAWHVMDTILSQSAPALHMAVRDLIRSVFLCDMDVFDAYL-PWRR 115
Db 59 GOLFGQHPDVFLMEPAWHVMDTILSQSAPALHMAVRDLIRAVFLCDMSVFDAYMEGPR 118

Qy 116 NISDLFQWAVSRALCSPPVCEAFARGNISSEBCKPLCATRPFGLAQACSSYSHVLKE 175
Db 119 RQSSLFQWENSALCSPACDIIIPQDEIIPRAHCRLLCSQQPFVEVEKACRSYSHVLKE 178

Qy 176 VRFFNLQVLYPLSDPALNLRIVHLVRDPAVLRVREQTAKALARDNGIVLGTNGTWV-E 234
Db 179 VRFFNLQVLYPLSDPALNLRIVHLVRDPAVLRVREQTAKALARDNGIVLGTNGTWV-E 238

Qy 235 ADPLRLVYVNEVCRSHVRIAEALHKPPFLQDRYLRVRYEDLARDPLTVIRELYAFTGLG 294
Db 235 ADPLRLVYVNEVCRSHVRIAEALHKPPFLQDRYLRVRYEDLARDPLTVIRELYAFTGLG 294

Qy 295 LTPQLQOTWIHNTHGSGPGARREAFKTTSRDALSVQAWRHHTLPFAKIRRVQELCGGALQ 354
Db 298 FLPLQOTWIHNTHGSGPGARREAFKTTSRDALSVQAWRHHTLPFAKIRRVQELCGGALQ 354

Qy 355 LLGYRSVHSELEQORDLSLDLL 375
Db 356 LLGYRHRVSESEQORNLLDLL 376

Query Match      49.5%; Score 1017.5; DB 13; Length 386;
Best Local Similarity 56.4%; Pred. No. 3e-92;
Matches 215; Conservative 43; Mismatches 112; Indels 11; Gaps 7;

Qy 1 MRLPRFSTVMSLLMVQTGILVF---LVSRQVPS-SPAGLGERVHVLVLSWRSGSFV 56
Db 1 MLLPK--KMKLLFLVSOQMAILALFFHMYSHNLSLWKQAQPERMHVVLVLSWRSGSFV 58

Qy 57 GOLFSQHPDVFLMEPAWHVMDTILSQSAPALHMAVRDLIRSVFLCDMDVFDAYL-PWRR 115
Db 59 GOLFGQHPDVFLMEPAWHVMDTILSQSAPALHMAVRDLIRAVFLCDMSVFDAYMEGPR 118

Qy 116 NISDLFQWAVSRALCSPPVCEAFARGNISSEBCKPLCATRPFGLAQACSSYSHVLKE 175
Db 119 RQSSLFQWENSALCSPACDIIIPQDEIIPRAHCRLLCSQQPFVEVEKACRSYSHVLKE 178

Qy 176 VRFFNLQVLYPLSDPALNLRIVHLVRDPAVLRVREQTAKALARDNGIVLGTNGTWV-E 234
Db 179 VRFFNLQVLYPLSDPALNLRIVHLVRDPAVLRVREQTAKALARDNGIVLGTNGTWV-E 238

Qy 235 ADPLRLVYVNEVCRSHVRIAEALHKPPFLQDRYLRVRYEDLARDPLTVIRELYAFTGLG 294
Db 235 ADPLRLVYVNEVCRSHVRIAEALHKPPFLQDRYLRVRYEDLARDPLTVIRELYAFTGLG 294

Qy 295 LTPQLQOTWIHNTHGSGPGARREAFKTTSRDALSVQAWRHHTLPFAKIRRVQELCGGALQ 354
Db 298 FLPLQOTWIHNTHGSGPGARREAFKTTSRDALSVQAWRHHTLPFAKIRRVQELCGGALQ 354

Qy 355 LLGYRSVHSELEQORDLSLDLL 375
Db 356 LLGYRHRVSESEQORNLLDLL 376

RESULT 14
US-10-841-707-6
; Sequence 6, Application US/10841707
; Publication No. US20040202649A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Minoru
```

```

; APPLICANT: Yeh, Jiunn-Chern
; APPLICANT: Hiraoka, Nobuyoshi
; TITLE OF INVENTION: Identification of the Meca-79 Antigen
; TITLE OF INVENTION: and Related Methods of Treating L-Selectin-Mediated
; TITLE OF INVENTION: Conditions
; FILE REFERENCE: P-LJ 4149
; CURRENT APPLICATION NUMBER: US/10/841,707
; CURRENT FILING DATE: 2004-05-06
; PRIOR APPLICATION NUMBER: US/09/569,320A
; PRIOR FILING DATE: 2000-05-11
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-841-707-6

Query Match      49.4%; Score 1015.5; DB 16; Length 380;
Best Local Similarity 57.1%; Pred. No. 4.6e-92;
Matches 212; Conservative 42; Mismatches 108; Indels 9; Gaps 6;

Qy 11 MSLMLVQTGILVF---LVSQVPS-SPAGLGERVHVLLVSSWSSGSGSFVCOLEFSQHPDV 66
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 3 LLLFLVSQMAILALFFHMYSHNLSLMSKQAPERMHVLLVSSWSSGSGSFVQLEFGQHPDV 62

Qy 67 FYLMEPAWHVMDTSLSQSAPALHMAVRDLIRSVFLCDMDVFDAYL-PWRRNISDLFQWAV 125
   ||||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 63 FYLMEPAWHVMTFKQSTAWMLHMAVRDLIRAVFLCDMSVFDAYWEPGPRQSSLFQWEN 122

Qy 126 SRALCSPPVCEAFARGNISSEVECKPLCATRPFGLAQEACSSYSHVVLKEVRFNQLQVLY 185
   ||||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 123 SRALCSAPACDIIIPQDEIIIPRAHCRLLCSQCFEVEVEKACRSYSHVVLKEVRFNQLQSLY 182

Qy 186 PLLSDPALNLRIVLVRDPRVAVLRSEQTAKALARDNGIVLGTNGTWV-EADPLRVVNE 244
   ||||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 183 PLLKDPSLNLHIVLVRDPRVAVFRSERTKGLMIDSRIVMGOHQEQLKKEDQPYVYMQV 242

Qy 245 VCRSHVRTAEALHKPPFPLODRVLRVRYEDLARDPLTVIRELYAFTGLTPOLOTWVH 304
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 243 ICQSQLEIYK-TIOSLPKALQERYLLVRYEDLARAPVAQTSRMVEFVGLFPLHLQTVWH 301

Qy 305 NITHGSGPGARREAFKTTSDALSVSQAWRHTLPFAKIRRVQELCGGALQLLGYRSVHSE 364
   ||||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 302 NITRGKMG--DHAFHTNARDALNVSOAWRWSLPYEKVSRLQKACGDAMNLLGYRHVRS 359

Qy 365 LEQRDLSLDLL 375
   ||||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 360 QEQRNLLLDLL 370

RESULT 15
US-10-427-631-11
; Sequence 11, Application US/10427631
; Publication No. US20030175923A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION; TANG, Y. Tom;
; APPLICANT: CORLEY, Neil C.; GUEGLER, Karl J.;
; APPLICANT: BAUGHN, Mariah R.; LAL, Preeti G.;
; APPLICANT: YUE, Henry; HILLMAN, Jennifer L.;
; APPLICANT: AZIMZAI, Yalda
; TITLE OF INVENTION: HUMAN TRANSFERASE PROTEINS
; FILE REFERENCE: PF-0592-1 DIV
; CURRENT APPLICATION NUMBER: US/10/427,631
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 09/786,240
; PRIOR FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: PCT/US99/20989
; PRIOR FILING DATE: 1999-09-09
; PRIOR APPLICATION NUMBER: US 60/172,220
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: US-60/155,248
; PRIOR FILING DATE: 1998-11-04

```

```

; PRIOR APPLICATION NUMBER: US 60/133,642
; PRIOR FILING DATE: 1999-05-11
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PERL Program
; SEQ ID NO 11
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030175923A1 2617407CD1
US-10-427-631-11

Query Match      48.0%; Score 986.5; DB 14; Length 386;
Best Local Similarity 56.3%; Pred. No. 3.6e-89;
Matches 215; Conservative 40; Mismatches 114; Indels 13; Gaps 9;

Qy 1 MRLPRSSSTVMSLLMVQTGILVF---LVSQVPS-SPAGLGERVHVLLVSSWSSGSGSFV 56
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 1 MLLLPK--KMKLLFLVSQMAILALFFHMYSHNLSLMSKQAPERMHVLLVSSWSSGSGSFV 58

Qy 57 GOLFSQHPDVFYLMPEPAWHVMDTSLSQSAPALHMAVRDLIRSVFLCDMDVFDAYL-PWRR 115
   ||||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 59 GOLFGQHPDVFYLMPEPAWHVMTFKQSTAWMLHMAVRDLIRAVFLCDMSVFDAYWEPGPR 118

Qy 116 NISDLFQWAVSRALCSPPVCEAFARGNISSEVECKPLCATR-PFGLAQEACSSYSHVVLK 174
   ||||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 119 RQSSLFQWENSRALCSAPACDIIIPQDE--SSPGLTAGSCAVNSPLKLEKACRSYSHVVLK 177

Qy 175 EVRFNQLQVLYPLLSDPALNLRIVLVRDPRVAVLRSEQTAKALARDNGIVLGTNGTWV- 233
   ||||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 178 EVRFNQLQSLYPLLKDPSLNLHIVLVRDPRVAVFRSERTKGLMIDSRIVMGOHQEQLK 237

Qy 234 EADPLRVVNEVCRSHVRTAEALHKPPFPLODRVLRVRYEDLARDPLTVIRELYAFTGL 293
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 238 KEDQPYVYMQVICQSQLEIYK-TIOSLPKALQERYLLVRYEDLARAPVAQTSRMVEFVGL 296

Qy 294 GLTPQLQTVIHNITHGSGPGARREAFKTTSDALSVSQAWRHTLPFAKIRRVQELCGGAL 353
   ||||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 297 EFLPHLQTVVHNITRGKMG--DHAFHTNARDALNVSOAWRWSLPYEKVSRLQKACGDAM 354

Qy 354 QLLGYRSVHSELEQRDLSLDLL 375
   ||||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 355 NLLGYRHVRSSEQEQRNLLLDLL 376

Search completed: June 23, 2005, 09:27:52
Job time : 51.935 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2005, 08:49:14 ; Search time 49.3029 Seconds
(without alignments)
3041.886 Million cell updates/sec

Title: US-10-697-828-8

Perfect score: 2051

Sequence: 1 MWLPRFSKTVTLVLAQT.....LTDLVLRGDFHFWASPD 390

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1717557 seqs, 384547976 residues

Total number of hits satisfying chosen parameters: 1717557

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published Applications AA:

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US10F_PUBCOMB.pep.*
- 19: /cgn2_6/ptodata/1/pubpaa/US11A_PUBCOMB.pep.*
- 20: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
- 21: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 22: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|---------------------|
| 1 | 2051 | 100.0 | 390 | 9 | US-09-927-602-4 |
| 2 | 2051 | 100.0 | 390 | 16 | US-10-697-828-8 |
| 3 | 1758.5 | 85.7 | 395 | 9 | US-09-927-602-3 |
| 4 | 1729.5 | 84.3 | 395 | 9 | US-09-927-602-2 |
| 5 | 1729.5 | 84.3 | 395 | 15 | US-10-258-080-5 |
| 6 | 1729.5 | 84.3 | 395 | 15 | US-10-648-593-159 |
| 7 | 1729.5 | 84.3 | 395 | 16 | US-10-697-828-13 |
| 8 | 1536.5 | 74.9 | 395 | 16 | US-10-697-828-7 |
| 9 | 1536.5 | 74.9 | 395 | 16 | US-10-841-707-8 |
| 10 | 1523 | 74.3 | 418 | 9 | US-09-927-602-5 |
| 11 | 1513.5 | 73.8 | 394 | 16 | US-10-408-765A-2211 |

| | | | | | | |
|----|--------|------|------|----|----------------------|--------------------|
| 12 | 1008 | 49.1 | 386 | 9 | US-09-816-825-2 | Sequence 2, Appli |
| 13 | 1008 | 49.1 | 386 | 13 | US-10-007-262-1 | Sequence 1, Appli |
| 14 | 1003.5 | 48.9 | 380 | 16 | US-10-841-707-6 | Sequence 6, Appli |
| 15 | 970 | 47.3 | 386 | 14 | US-10-427-631-11 | Sequence 11, Appli |
| 16 | 808 | 39.4 | 169 | 9 | US-09-927-602-7 | Sequence 7, Appli |
| 17 | 755 | 36.8 | 169 | 9 | US-09-927-602-6 | Sequence 6, Appli |
| 18 | 665 | 32.4 | 483 | 14 | US-10-212-933-2 | Sequence 2, Appli |
| 19 | 659.5 | 32.2 | 530 | 16 | US-10-723-860-1409 | Sequence 1409, Ap |
| 20 | 656.5 | 32.0 | 484 | 14 | US-10-212-933-4 | Sequence 4, Appli |
| 21 | 656.5 | 32.0 | 531 | 9 | US-09-833-790-255 | Sequence 255, App |
| 22 | 656.5 | 32.0 | 531 | 16 | US-10-755-889-284 | Sequence 284, App |
| 23 | 579.5 | 28.3 | 479 | 13 | US-10-087-192-126 | Sequence 126, App |
| 24 | 576 | 28.1 | 481 | 13 | US-10-087-192-123 | Sequence 123, App |
| 25 | 514.5 | 25.1 | 171 | 9 | US-09-927-602-8 | Sequence 8, Appli |
| 26 | 513.5 | 25.0 | 411 | 14 | US-10-021-660-128 | Sequence 128, App |
| 27 | 513.5 | 25.0 | 411 | 15 | US-10-211-462-97 | Sequence 97, Appl |
| 28 | 513.5 | 25.0 | 411 | 16 | US-10-408-765A-395 | Sequence 395, App |
| 29 | 513.5 | 25.0 | 411 | 16 | US-10-723-860-1544 | Sequence 1544, Ap |
| 30 | 372 | 18.1 | 169 | 9 | US-09-927-602-9 | Sequence 9, Appli |
| 31 | 351.5 | 17.1 | 174 | 9 | US-09-927-602-11 | Sequence 11, Appl |
| 32 | 335 | 16.3 | 179 | 9 | US-09-927-602-10 | Sequence 10, Appl |
| 33 | 119.5 | 5.8 | 596 | 16 | US-10-697-828-9 | Sequence 9, Appli |
| 34 | 119.5 | 5.8 | 1212 | 17 | US-10-479-472A-2 | Sequence 2, Appli |
| 35 | 119.5 | 5.8 | 1222 | 16 | US-10-697-828-15 | Sequence 15, Appl |
| 36 | 119.5 | 5.8 | 1222 | 16 | US-10-475-446-4 | Sequence 4, Appli |
| 37 | 112.5 | 5.5 | 1207 | 16 | US-10-697-828-17 | Sequence 17, Appl |
| 38 | 112.5 | 5.5 | 1207 | 17 | US-10-479-472A-4 | Sequence 4, Appli |
| 39 | 105 | 5.1 | 359 | 14 | US-10-411-976-11 | Sequence 11, Appl |
| 40 | 102.5 | 5.0 | 387 | 14 | US-10-126-279-21 | Sequence 21, Appl |
| 41 | 102.5 | 5.0 | 387 | 14 | US-10-286-606-21 | Sequence 21, Appl |
| 42 | 102.5 | 5.0 | 387 | 16 | US-10-891-383-21 | Sequence 21, Appl |
| 43 | 97 | 4.7 | 299 | 14 | US-10-156-761-9846 | Sequence 9846, Ap |
| 44 | 95 | 4.6 | 885 | 17 | US-10-732-923-6830 | Sequence 6830, Ap |
| 45 | 94.5 | 4.6 | 2588 | 16 | US-10-437-963-138279 | Sequence 138279, |

ALIGNMENTS

RESULT 1
US-09-927-602-4
; Sequence 4, Application US/09927602
; Patent No. US20020061562A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; APPLICANT: Akama, Tomoya O.
; TITLE OF INVENTION: Methods of Treating Macular Corneal
; FILE REFERENCE: P-LJ 4852
; CURRENT FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 09/638,211
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 390
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-927-602-4

| | | | | |
|-----------------------|-----------------|--|-----------|-------------|
| Query Match | 100.0% | Score 2051; | DB 9; | Length 390; |
| Best Local Similarity | 100.0%; | Pred. No. 2.4e-204; | | |
| Matches 390; | Conservative 0; | Mismatches 0; | Indels 0; | Gaps 0; |
| Qy | 1 | MWLPFRFSKTVTLVLAQTCLLLFIISRRGPPSPAGGEDRVHVLVSSWFGSFLGQL | 60 | |
| Db | 1 | MWLPFRFSKTVTLVLAQTCLLLFIISRRGPPSPAGGEDRVHVLVSSWFGSFLGQL | 60 | |
| Qy | 61 | FSQHPDVFYLMERPAHWVTTLSQGSAAATLHMAVRDLMRSLFLCDMDVFDAYMPSRNL5A | 120 | |
| Db | 61 | FSQHPDVFYLMERPAHWVTTLSQGSAAATLHMAVRDLMRSLFLCDMDVFDAYMPSRNL5A | 120 | |

QY 121 FFWATSRALCSPACSAFPRGTISKQDVCKTLCTROPFSLAREACRSYSHVVYLKEVREF 180
DB 121 FFWATSRALCSPACSAFPRGTISKQDVCKTLCTROPFSLAREACRSYSHVVYLKEVREF 180
QY 181 NLQVLYPLSDPALNLRIHVLRDPRAVLRSRRAAGPILARDNGIVLGTNGKWEADPHL 240
DB 181 NLQVLYPLSDPALNLRIHVLRDPRAVLRSRRAAGPILARDNGIVLGTNGKWEADPHL 240
QY 241 RLIREVCRSHVRIAEATLKPPFLRGYRLVRFEDLAREPLAEIRALYAFTGLTLPOL 300
DB 241 RLIREVCRSHVRIAEATLKPPFLRGYRLVRFEDLAREPLAEIRALYAFTGLTLPOL 300
QY 301 EAWIHNTHSGGIGKPIEAFTSSRNARNVQAWRHALPFTKILRQVEVCAGALQLLGYR 360
DB 301 EAWIHNTHSGGIGKPIEAFTSSRNARNVQAWRHALPFTKILRQVEVCAGALQLLGYR 360
QY 361 PVSADQQRDLTDLVLPRGPDHFSWASPD 390
DB 361 PVSADQQRDLTDLVLPRGPDHFSWASPD 390
RESULT 2
US-10-697-828-8
; Sequence 8, Application US/10697828
; Publication No. US20040185546A1
; GENERAL INFORMATION:
; APPLICANT: Rosen, Steven
; APPLICANT: Lee, Jin Kyu
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: Novel Glycosyl Sulfotransferases GST-4alpha, GST-4beta, & GST-6
; FILE REFERENCE: UCAL-138DIV
; CURRENT APPLICATION NUMBER: US/10/697,828
; CURRENT FILING DATE: 2003-10-29
; PRIOR APPLICATION NUMBER: 09/593,828
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: 60/144,694
; PRIOR FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 390
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-697-828-8
Query Match 100.0%; Score 2051; DB 16; Length 390;
Best Local Similarity 100.0%; Pred. No. 2.4e-204;
Matches 390; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MWLPRFSSKTVTVLLLAQTTCLLLFIISRPSPSPAGGEDRVHVLVLSWRSGSFLGOL 60
DB 1 MWLPRFSSKTVTVLLLAQTTCLLLFIISRPSPSPAGGEDRVHVLVLSWRSGSFLGOL 60
QY 61 FSQHPDVFLMEPAWHVMTTLSQGSAAATLHMAVRDLMRSLFLCDMDVFDAYMPSQSNLSA 120
DB 61 FSQHPDVFLMEPAWHVMTTLSQGSAAATLHMAVRDLMRSLFLCDMDVFDAYMPSQSNLSA 120
QY 121 FFWATSRALCSPACSAFPRGTISKQDVCKTLCTROPFSLAREACRSYSHVVYLKEVREF 180
DB 121 FFWATSRALCSPACSAFPRGTISKQDVCKTLCTROPFSLAREACRSYSHVVYLKEVREF 180
QY 181 NLQVLYPLSDPALNLRIHVLRDPRAVLRSRRAAGPILARDNGIVLGTNGKWEADPHL 240
DB 181 NLQVLYPLSDPALNLRIHVLRDPRAVLRSRRAAGPILARDNGIVLGTNGKWEADPHL 240
QY 241 RLIREVCRSHVRIAEATLKPPFLRGYRLVRFEDLAREPLAEIRALYAFTGLTLPOL 300
DB 241 RLIREVCRSHVRIAEATLKPPFLRGYRLVRFEDLAREPLAEIRALYAFTGLTLPOL 300
QY 301 EAWIHNTHSGGIGKPIEAFTSSRNARNVQAWRHALPFTKILRQVEVCAGALQLLGYR 360
DB 301 EAWIHNTHSGGIGKPIEAFTSSRNARNVQAWRHALPFTKILRQVEVCAGALQLLGYR 360

QY 361 PVSADQQRDLTDLVLPRGPDHFSWASPD 390
DB 361 PVSADQQRDLTDLVLPRGPDHFSWASPD 390
RESULT 3
US-09-927-602-3
; Sequence 3, Application US/09927602
; Patent No. US20020061562A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; APPLICANT: Akama, Tomoya O.
; TITLE OF INVENTION: Methods of Treating Macular Corneal
; TITLE OF INVENTION: Dystrophy
; FILE REFERENCE: P-LJ 4852
; CURRENT APPLICATION NUMBER: US/09/927,602
; CURRENT FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 09/638,211
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: synthetic construct
; NAME/KEY: VARIANT
; LOCATION: (1)...(395)
; OTHER INFORMATION: Xaa = any amino acid
US-09-927-602-3
Query Match 85.7%; Score 1758.5; DB 9; Length 395;
Best Local Similarity 87.1%; Pred. No. 6.9e-174;
Matches 338; Conservative 10; Mismatches 39; Indels 1; Gaps 1;
QY 1 MWLPRFSSKTVTVLLLAQTTCLLLFIISRPSPSPAGGEDRVHVLVLSWRSGSFLGOL 60
DB 1 MWLPRFSSKTVTVLLLAQTTCLLLFIISRPSPSPAGGEDRVHVLVLSWRSGSFLGOL 59
QY 61 FSQHPDVFLMEPAWHVMTTLSQGSAAATLHMAVRDLMRSLFLCDMDVFDAYMPSQSNLSA 120
DB 61 FSQHPDVFLMEPAWHVMTTLSQGSAAATLHMAVRDLMRSLFLCDMDVFDAYMPSQSNLSA 119
QY 121 FFWATSRALCSPACSAFPRGTISKQDVCKTLCTROPFSLAREACRSYSHVVYLKEVREF 180
DB 121 FFWATSRALCSPACSAFPRGTISKQDVCKTLCTROPFSLAREACRSYSHVVYLKEVREF 179
QY 181 NLQVLYPLSDPALNLRIHVLRDPRAVLRSRRAAGPILARDNGIVLGTNGKWEADPHL 240
DB 181 NLQVLYPLSDPALNLRIHVLRDPRAVLRSRRAAGPILARDNGIVLGTNGKWEADPHL 239
QY 241 RLIREVCRSHVRIAEATLKPPFLRGYRLVRFEDLAREPLAEIRALYAFTGLTLPOL 300
DB 241 RLIREVCRSHVRIAEATLKPPFLRGYRLVRFEDLAREPLAEIRALYAFTGLTLPOL 299
QY 301 EAWIHNTHSGGIGKPIEAFTSSRNARNVQAWRHALPFTKILRQVEVCAGALQLLGYR 360
DB 301 EAWIHNTHSGGIGKPIEAFTSSRNARNVQAWRHALPFTKILRQVEVCAGALQLLGYR 359
QY 361 PVSADQQRDLTDLVLPRGPDHFSWAS 388
DB 361 PVSADQQRDLTDLVLPRGPDHFSWAS 387
RESULT 4
US-09-927-602-2
; Sequence 2, Application US/09927602
; Patent No. US20020061562A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; APPLICANT: Akama, Tomoya O.
; TITLE OF INVENTION: Methods of Treating Macular Corneal

; TITLE OF INVENTION: Dystrophy
 ; FILE REFERENCE: P-LJ 4852
 ; CURRENT APPLICATION NUMBER: US/09/927.602
 ; CURRENT FILING DATE: 2001-08-09
 ; PRIOR APPLICATION NUMBER: US 09/638,211
 ; PRIOR FILING DATE: 2000-08-11
 ; NUMBER OF SEQ ID NOS: 38
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 2
 ; LENGTH: 395
 ; TYPE: PRT
 ; ORGANISM: Homo Sapien
 ; US-09-927-602-2

Query Match 84.3%; Score 1729.5; DB 9; Length 395;
 Best Local Similarity 85.8%; Pred. No. 7.2e-171;
 Matches 333; Conservative 17; Mismatches 37; Indels 1; Gaps 1;
 QY 1 MWLPRFSKTVTVLLAQTCLLFIISRPSPSPAGGEDRVHVLVLSWRSRSGSFLGQL 60
 DB 1 MWLPRVSSSTAVTALLAQ-TFLLFLVSRPGSPSPAGGEARVHVLVLSWRSRSGSFLGQL 59
 QY 61 FSOHPDVFLMEPAWHVWTTLSQSAATLHMVARDLMRSIFLCMDVDFDAYMPOSRLSA 120
 DB 60 FNOHPDVFLMEPAWHVWTTLSQSAATLHMVARDLMRSIFLCMDVDFDAYLPMWRRLSD 119
 QY 121 FFWNATSRALCSPSPACSAFPRGTISKQDVCKTCTROPFSLAREACRSYSHVVLKEVRF 180
 DB 120 LFQWAVSRALCSPSPACSAFPRGAISSAECVKPLCARQSFILAREACRSYSHVVLKEVRF 179
 QY 181 NLQVLYPLSDPALNLRIHVLRDPRAVLRSREAAAGPILARDNGIVLGTNGKWEADPHL 240
 DB 180 NLQVLYPLSDPALNLRIHVLRDPRAVLRSREAAAGPILARDNGIVLGTNGKWEADPHL 239
 QY 241 RLIREVCRSHVRIAEATLKPPPLRGYRLVRFDLAREPLAEIRALYAFGLTLPOL 300
 DB 240 RVREVCRRSHVRIAEATLKPPPLRGYRLVRFDLAREPLAEIRALYAFGLTLPOL 299
 QY 301 EAWIHNTHGSGIGKPIEAFTSSRNARNVSOQWRHALPFTKILRVOECAGALQLLGYR 360
 DB 300 EAWIHNTHGSGIGKPIEAFTSSRNARNVSOQWRHALPFTKILRVOECAGALQLLGYR 359
 QY 361 PVYSADQQRDLTLDLVLPRGPDHFSWAS 388
 DB 360 PVYSEDEQRNLALDLVLPRLGNGFTWAS 387

RESULT 5
 US-10-258-080-5
 ; Sequence 5, Application US/10258080
 ; Publication No. US20040029125A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Incyte Genomics, Inc.
 ; APPLICANT: POLICKY, Jennifer L.
 ; APPLICANT: HAPALIA, April J.A.
 ; APPLICANT: BURFORD, Neil
 ; APPLICANT: RING, Huijun Z.
 ; APPLICANT: LAL, Freeti G.
 ; APPLICANT: TRIBOULEY, Catherine M.
 ; APPLICANT: YAO, Monique G.
 ; APPLICANT: YUE, Henry
 ; APPLICANT: TANG, Y. Tom
 ; APPLICANT: ARVIZU, Chandra S.
 ; APPLICANT: DAS, Debopriya
 ; APPLICANT: SANJANWALA, Madhusudan M.
 ; APPLICANT: GANDHI, Aneena R.
 ; APPLICANT: REDDY, Roopa M.
 ; APPLICANT: KHAN, Farrah A.
 ; APPLICANT: BAUGHN, Mariah R.
 ; APPLICANT: RAMKUMAR, Jayalaxmi
 ; APPLICANT: GRIFFIN, Jennifer A.
 ; APPLICANT: AU-YOUNG, Janice K.
 ; TITLE OF INVENTION: DRUG METABOLIZING ENZYMES

; FILE REFERENCE: PI-0070 USN
 ; CURRENT APPLICATION NUMBER: US/10/258,080
 ; CURRENT FILING DATE: 2002-10-15
 ; PRIOR APPLICATION NUMBER: US 60/203,509
 ; PRIOR FILING DATE: 2000-05-11
 ; PRIOR APPLICATION NUMBER: US 60/202,234
 ; PRIOR FILING DATE: 2000-05-05
 ; PRIOR APPLICATION NUMBER: US 60/200,185
 ; PRIOR FILING DATE: 2000-04-28
 ; PRIOR APPLICATION NUMBER: US 60/198,403
 ; PRIOR FILING DATE: 2000-04-19
 ; PRIOR APPLICATION NUMBER: PCT/US01/11869
 ; PRIOR FILING DATE: 2001-04-12
 ; PRIOR APPLICATION NUMBER: US 60/197,590
 ; PRIOR FILING DATE: 2000-04-13
 ; NUMBER OF SEQ ID NOS: 20
 ; SOFTWARE: PERL Program
 ; SEQ ID NO 5
 ; LENGTH: 395
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; OTHER INFORMATION: Incyte ID No. US20040029125A1 7472777CD1
 ; US-10-258-080-5

Query Match 84.3%; Score 1729.5; DB 15; Length 395;
 Best Local Similarity 85.8%; Pred. No. 7.2e-171;
 Matches 333; Conservative 17; Mismatches 37; Indels 1; Gaps 1;
 QY 1 MWLPRFSKTVTVLLAQTCLLFIISRPSPSPAGGEDRVHVLVLSWRSRSGSFLGQL 60
 DB 1 MWLPRVSSSTAVTALLAQ-TFLLFLVSRPGSPSPAGGEARVHVLVLSWRSRSGSFLGQL 59
 QY 61 FSOHPDVFLMEPAWHVWTTLSQSAATLHMVARDLMRSIFLCMDVDFDAYMPOSRLSA 120
 DB 60 FNOHPDVFLMEPAWHVWTTLSQSAATLHMVARDLMRSIFLCMDVDFDAYLPMWRRLSD 119
 QY 121 FFWNATSRALCSPSPACSAFPRGTISKQDVCKTCTROPFSLAREACRSYSHVVLKEVRF 180
 DB 120 LFQWAVSRALCSPSPACSAFPRGAISSAECVKPLCARQSFILAREACRSYSHVVLKEVRF 179
 QY 181 NLQVLYPLSDPALNLRIHVLRDPRAVLRSREAAAGPILARDNGIVLGTNGKWEADPHL 240
 DB 180 NLQVLYPLSDPALNLRIHVLRDPRAVLRSREAAAGPILARDNGIVLGTNGKWEADPHL 239
 QY 241 RLIREVCRSHVRIAEATLKPPPLRGYRLVRFDLAREPLAEIRALYAFGLTLPOL 300
 DB 240 RVREVCRRSHVRIAEATLKPPPLRGYRLVRFDLAREPLAEIRALYAFGLTLPOL 299
 QY 301 EAWIHNTHGSGIGKPIEAFTSSRNARNVSOQWRHALPFTKILRVOECAGALQLLGYR 360
 DB 300 EAWIHNTHGSGIGKPIEAFTSSRNARNVSOQWRHALPFTKILRVOECAGALQLLGYR 359
 QY 361 PVYSADQQRDLTLDLVLPRGPDHFSWAS 388
 DB 360 PVYSEDEQRNLALDLVLPRLGNGFTWAS 387

RESULT 6
 US-10-648-593-159
 ; Sequence 159, Application US/10648593
 ; Publication No. US20040106132A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Bristol-Myers Squibb Company
 ; TITLE OF INVENTION: IDENTIFICATION OF GENES FOR PREDICTING ACTIVITY OF COMPOUNDS THAT
 ; INTERACT WITH AND/OR MODULATE PROTEIN TYROSINE KINASES AND/OR
 ; TITLE OF INVENTION: INTERACT WITH AND/OR MODULATE PROTEIN TYROSINE KINASES AND/OR
 ; TITLE OF INVENTION: PROTEIN TYROSINE KINASE PATHWAYS IN BREAST CELLS
 ; FILE REFERENCE: D0273 NP
 ; CURRENT APPLICATION NUMBER: US/10/648,593
 ; CURRENT FILING DATE: 2003-08-26
 ; PRIOR APPLICATION NUMBER: 60/406,385
 ; PRIOR FILING DATE: 2002-08-27

; NUMBER OF SEQ ID NOS: 557
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 159
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-648-593-159

Query Match 84.3%; Score 1729.5; DB 16; Length 395;
Best Local Similarity 85.8%; Pred. No. 7.2e-171;
Matches 333; Conservative 17; Mismatches 37; Indels 1; Gaps 1;

Qy 1 MWLPRFSSKTVTVLLAQTCLLFIISRPGPSAGGEDRVHVLVSSWRSGSSFLGQL 60
Db 1 MWLPRVSVSTAVTALLAQ-TFLLLFIVSRPGPSPAGGEARVHVLVSSWRSGSSFLGQL 59
Qy 61 FSQHPDVFYLMPEPAWHVWTTLSQGSAAATLHMAVRDLMSIFLCMDVDFDAYMPOSRNLSA 120
Db 60 FNOHPDVFYLMPEPAWHVWTTLSQGSAAATLHMAVRDLVRSVFLCMDVDFDAYLPWRNLS 119
Qy 121 FFWNATSRALCSPACSAFPRGTISKQDVCKTLCTROPFSLAREACRSYSHVVVKVRRFF 180
Db 120 LFQWAVSRALCSPACSAFPRGAISSEAVCKPLCARQSFTLAREACRSYSHVVVKVRRFF 179
Qy 181 NLQVLYPLLSDPALNLRIVHLVRDPRAVLRSEAAAGPILARDNGIVLGTNGKWEADPHL 240
Db 180 NLQVLYPLLSDPALNLRIVHLVRDPRAVLRSEQTAKALARDNGIVLGTNGTWVEADPGL 239
Qy 241 RLIREVCRSHVRIAEAAATLKPPPLRGRVRLVRFEDLAREPLAEIRALYAFTGLTLPQL 300
Db 240 RVREVCVSHVRIAEAAATLKPPPLRGRVRLVRFEDLAREPLAEIRALYAFTGLTLPQL 299
Qy 301 EAWIHNTGSGIGKPIEAFTSSRNARNVSQAWRHALPFTKILRVOECAGALQLLGYR 360
Db 300 EAWIHNTGSGPGARREAFKTSRNALNVSQAWRHALPFAKIRRVQOELCAGALQLLGYR 359
Qy 361 PVSADQORDLTLDLVLPRGPDHFSWAS 388
Db 360 PVSDEQORNALDLVLPRGLNGFTWAS 387

RESULT 7

US-10-697-828-13
; Sequence 13, Application US/10697828
; Publication No. US20040185546A1
; GENERAL INFORMATION:
; APPLICANT: Rosen, Steven
; APPLICANT: Lee, Jin Kyu
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: Novel Glycosyl Sulfotransferases GST-4alpha, GST-4beta, & GST-6
; FILE REFERENCE: UCAL-138DIV
; CURRENT APPLICATION NUMBER: US/10/697,828
; PRIOR FILING DATE: 2003-10-29
; PRIOR APPLICATION NUMBER: 09/593,828
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: 60/144,694
; PRIOR FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 395
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-697-828-13

Query Match 84.3%; Score 1729.5; DB 16; Length 395;
Best Local Similarity 85.8%; Pred. No. 7.2e-171;
Matches 333; Conservative 17; Mismatches 37; Indels 1; Gaps 1;

Qy 1 MWLPRFSSKTVTVLLAQTCLLFIISRPGPSAGGEDRVHVLVSSWRSGSSFLGQL 60
Db 1 MWLPRVSVSTAVTALLAQ-TFLLLFIVSRPGPSPAGGEARVHVLVSSWRSGSSFLGQL 59

Qy 61 FSQHPDVFYLMPEPAWHVWTTLSQGSAAATLHMAVRDLMSIFLCMDVDFDAYMPOSRNLSA 120
Db 60 FNOHPDVFYLMPEPAWHVWTTLSQGSAAATLHMAVRDLVRSVFLCMDVDFDAYLPWRNLS 119
Qy 121 FFWNATSRALCSPACSAFPRGTISKQDVCKTLCTROPFSLAREACRSYSHVVVKVRRFF 180
Db 120 LFQWAVSRALCSPACSAFPRGAISSEAVCKPLCARQSFTLAREACRSYSHVVVKVRRFF 179
Qy 181 NLQVLYPLLSDPALNLRIVHLVRDPRAVLRSEAAAGPILARDNGIVLGTNGKWEADPHL 240
Db 180 NLQVLYPLLSDPALNLRIVHLVRDPRAVLRSEQTAKALARDNGIVLGTNGTWVEADPGL 239
Qy 241 RLIREVCRSHVRIAEAAATLKPPPLRGRVRLVRFEDLAREPLAEIRALYAFTGLTLPQL 300
Db 240 RVREVCVSHVRIAEAAATLKPPPLRGRVRLVRFEDLAREPLAEIRALYAFTGLTLPQL 299
Qy 301 EAWIHNTGSGIGKPIEAFTSSRNARNVSQAWRHALPFTKILRVOECAGALQLLGYR 360
Db 300 EAWIHNTGSGPGARREAFKTSRNALNVSQAWRHALPFAKIRRVQOELCAGALQLLGYR 359
Qy 361 PVSADQORDLTLDLVLPRGPDHFSWAS 388
Db 360 PVSDEQORNALDLVLPRGLNGFTWAS 387

RESULT 8

US-10-697-828-7
; Sequence 7, Application US/10697828
; Publication No. US20040185546A1
; GENERAL INFORMATION:
; APPLICANT: Rosen, Steven
; APPLICANT: Lee, Jin Kyu
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: Novel Glycosyl Sulfotransferases GST-4alpha, GST-4beta, & GST-6
; FILE REFERENCE: UCAL-138DIV
; CURRENT APPLICATION NUMBER: US/10/697,828
; PRIOR FILING DATE: 2003-10-29
; PRIOR APPLICATION NUMBER: 09/593,828
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: 60/144,694
; PRIOR FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 395
; TYPE: PRT
; ORGANISM: mus musculus
US-10-697-828-7

Query Match 74.9%; Score 1536.5; DB 16; Length 395;
Best Local Similarity 76.0%; Pred. No. 8.9e-151;
Matches 295; Conservative 30; Mismatches 62; Indels 1; Gaps 1;

Qy 1 MWLPRFSSKTVTVLLAQTCLLFIISRPGPSAGGEDRVHVLVSSWRSGSSFLGQL 60
Db 1 MWLPRVSVTVMLSLMVQTG-ILVFLVSRQVPSPPAGLGERVHVLVSSWRSGSSFLGQL 59
Qy 61 FSQHPDVFYLMPEPAWHVWTTLSQGSAAATLHMAVRDLMSIFLCMDVDFDAYMPOSRNLSA 120
Db 60 FSQHPDVFYLMPEPAWHVWDTLSQGSAPALHMAVRDLIRSVFLCMDVDFDAYLPWRNLS 119
Qy 121 FFWNATSRALCSPACSAFPRGTISKQDVCKTLCTROPFSLAREACRSYSHVVVKVRRFF 180
Db 120 LFQWAVSRALCSPFPVCEAFARGNISSEEVCKPLCATRPFGLAQACSSYSHVVVKVRRFF 179
Qy 181 NLQVLYPLLSDPALNLRIVHLVRDPRAVLRSEAAAGPILARDNGIVLGTNGKWEADPHL 240
Db 180 NLQVLYPLLSDPALNLRIVHLVRDPRAVLRSEQTAKALARDNGIVLGTNGTWVEADPRL 239
Qy 241 RLIREVCRSHVRIAEAAATLKPPPLRGRVRLVRFEDLAREPLAEIRALYAFTGLTLPQL 300
Db 240 RVREVCVSHVRIAEAAALHKPPPLQDRVRLVRYEDLARDPLTVIRLYAFTGLTLPQL 299

QY 301 EAMHNTHGSGIGKPIEAFTSSRNARNVSQAWRHLPFTKILRVOECAGALQLLGYR 360
 Db 300 QTWIHNTHGSGPGARREAFKTTSRDALSVSQAWRHTLPFAKIRRVQELCGGALQLLGYR 359
 QY 361 PVYSADOQDRLTLVLPRGDPHFSWAS 388
 Db 360 SVHSELEQRDLSDLLPLRGMDSPFKWAS 387

RESULT 9

US-10-841-707-8
 ; Sequence 8, Application US/10841707
 ; Publication No. US20040202649A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Fukuda, Minoru
 ; APPLICANT: Yeh, Jiunn-Chern
 ; APPLICANT: Hiraoka, Nobuyoshi
 ; TITLE OF INVENTION: Identification of the Meca-79 Antigen
 ; TITLE OF INVENTION: and Related Methods of Treating L-Selectin-Mediated
 ; TITLE OF INVENTION: Conditions
 ; FILE REFERENCE: P-LJ 4149
 ; CURRENT APPLICATION NUMBER: US/10/841,707
 ; CURRENT FILING DATE: 2004-05-06
 ; PRIOR APPLICATION NUMBER: US/09/569,320A
 ; PRIOR FILING DATE: 2000-05-11
 ; NUMBER OF SEQ ID NOS: 29
 ; SOFTWARE: FastSEQ for Windows Version 4.0
 ; SEQ ID NO 8
 ; LENGTH: 395
 ; TYPE: PRT
 ; ORGANISM: Mus musculus
 US-10-841-707-8

Query Match 74.9%; Score 1536.5; DB 16; Length 395;
 Best Local Similarity 76.0%; Pred. No. 8.9e-151;
 Matches 295; Conservative 30; Mismatches 62; Indels 1; Gaps 1;

QY 1 MMLPRFSSKTVTLVLLAQTCLLLFIISRPSPAGGEDRVHVLVLSWRSFGSLGQL 60
 Db 1 MMLPRFSSKTVTLVLLAQTCLLLFIISRPSPAGGEDRVHVLVLSWRSFGSLGQL 59
 QY 61 FSOHPDVFYLMPEAWHVTTLTSGSAATLHMVDRDLMSIFLCMDVDFDAYMPOSRLSA 120
 Db 60 FSOHPDVFYLMPEAWHVTTLTSGSAATLHMVDRDLMSIFLCMDVDFDAYMPOSRLSA 119
 QY 121 FFWNATSRALCSPACAPPRGTISKQDVCKTLCTROPFSLAREACRSYSHVVLKEVRF 180
 Db 120 LFWNATSRALCSPACAPPRGTISKQDVCKTLCTROPFSLAREACRSYSHVVLKEVRF 179
 QY 181 NLQVLYPLSDPALNLRIVHLVRDPRAVLRSREAGPILARDNGIVLTNGKWEADPHL 240
 Db 180 NLQVLYPLSDPALNLRIVHLVRDPRAVLRSREAGPILARDNGIVLTNGKWEADPHL 239
 QY 241 RLIREVCRSHVRIAEATLKPFPPLGRYRLVRFEDLAREPLAEIRALYFTGLTLPQL 300
 Db 240 RVNEVCRSHVRIAEALHKKPPFLQDRYRLVRYEDLARDPLTVIRELYAFTGLTLPQL 299
 QY 301 EAMHNTHGSGIGKPIEAFTSSRNARNVSQAWRHLPFTKILRVOECAGALQLLGYR 360
 Db 300 QTWIHNTHGSGPGARREAFKTTSRDALSVSQAWRHTLPFAKIRRVQELCGGALQLLGYR 359
 QY 361 PVYSADOQDRLTLVLPRGDPHFSWAS 388
 Db 360 SVHSELEQRDLSDLLPLRGMDSPFKWAS 387

RESULT 10

US-09-927-602-5
 ; Sequence 5, Application US/09927602
 ; Patent No. US2002006152A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Fukuda, Michiko N.
 ; APPLICANT: Akama, Tomoya O.

; TITLE OF INVENTION: Methods of Treating Macular Corneal
 ; TITLE OF INVENTION: Dystrophy
 ; FILE REFERENCE: P-LJ 4852
 ; CURRENT APPLICATION NUMBER: US/09/927,602
 ; CURRENT FILING DATE: 2001-08-09
 ; PRIOR APPLICATION NUMBER: US 09/638,211
 ; PRIOR FILING DATE: 2000-08-11
 ; NUMBER OF SEQ ID NOS: 38
 ; SOFTWARE: FastSEQ for Windows Version 4.0
 ; SEQ ID NO 5
 ; LENGTH: 418
 ; TYPE: PRT
 ; ORGANISM: Mus musculus
 US-09-927-602-5

Query Match 74.3%; Score 1523; DB 9; Length 418;
 Best Local Similarity 75.8%; Pred. No. 2.5e-149;
 Matches 294; Conservative 30; Mismatches 62; Indels 2; Gaps 2;
 QY 1 MMLPRFSSKTVTLVLLAQTCLLLFIISRPSPAGGEDRVHVLVLSWRSFGSLGQL 60
 Db 25 MMLPRFSSKTVTLVLLAQTCLLLFIISRPSPAGGEDRVHVLVLSWRSFGSLGQL 83
 QY 61 FSOHPDVFYLMPEAWHVTTLTSGSAATLHMVDRDLMSIFLCMDVDFDAYMPOSRLSA 120
 Db 84 FSOHPDVFYLMPEAWHVTTLTSGSAATLHMVDRDLMSIFLCMDVDFDAYMPOSRLSA 143
 QY 121 FFWNATSRALCSPACAPPRGTISKQDVCKTLCTROPFSLAREACRSYSHVVLKEVRF 180
 Db 144 LFWNATSRALCSPACAPPRGTISKQDVCKTLCTROPFSLAREACRSYSHVVLKEVRF 203
 QY 181 NLQVLYPLSDPALNLRIVHLVRDPRAVLRSREAGPILARDNGIVLTNGKWEADPHL 240
 Db 204 NLQVLYPLSDPALNLRIVHLVRDPRAVLRSREAGPILARDNGIVLTNGKWEADPHL 263
 QY 241 RLIREVCRSHVRIAEATLKPFPPLGRYRLVRFEDLAREPLAEIRALYFTGLTLPQL 300
 Db 264 RVNEVCRSHVRIAEALHKKPPFLQDRYRLVRYEDLARDPLTVIRELYAFTGLTLPQL 322
 QY 301 EAMHNTHGSGIGKPIEAFTSSRNARNVSQAWRHLPFTKILRVOECAGALQLLGYR 360
 Db 323 QTWIHNTHGSGPGARREAFKTTSRDALSVSQAWRHTLPFAKIRRVQELCGGALQLLGYR 382
 QY 361 PVYSADOQDRLTLVLPRGDPHFSWAS 388
 Db 383 SVHSELEQRDLSDLLPLRGMDSPFKWAS 410

RESULT 11

US-10-408-765A-2211
 ; Sequence 2211, Application US/10408765A
 ; Publication No. US2004010187A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Ghosh, Soumitra S.
 ; APPLICANT: Fahy, Eoin D.
 ; APPLICANT: Zhang, Bing
 ; APPLICANT: Gibson, Bradford W.
 ; APPLICANT: Taylor, Steven W.
 ; APPLICANT: Glenn, Gary M.
 ; APPLICANT: Warnock, Dale E.
 ; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
 ; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
 ; FILE REFERENCE: 660088.465
 ; CURRENT APPLICATION NUMBER: US/10/408,765A
 ; CURRENT FILING DATE: 2003-04-04
 ; NUMBER OF SEQ ID NOS: 3077
 ; SOFTWARE: FastSEQ for Windows Version 4.0
 ; SEQ ID NO 2211
 ; LENGTH: 394
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-408-765A-2211

```
Query Match 73.8%; Score 1513.5; DB 16; Length 394;
Best Local Similarity 92.6%; Pred. No. 2.2e-148;
Matches 289; Conservative 7; Mismatches 15; Indels 1; Gaps 1;

Qy 1 MWLPRFSKVTYVULLAQTCLLLFIISRCPPSPAGGEDRVHVLVLSWRSGSSFLGQL 60
Db 1 MWLPRVSTAVTALLAQ-TFLLLFLVSRCPSPAGGEARVHVLVLSWRSGSSFLGQL 59
Qy 61 FSQHPDVFYLMPEAWHVTTLTSGSAATLHMAVRDLMSIFLDCMDVDFDYMPOSRNLSA 120
Db 60 FQHPDVFYLMPEAWHVTTLTSGSAATLHMAVRDLMSIFLDCMDVDFDYLPRNRNLS 119
Qy 121 FFWNATSRALCSPACSAFPRGTISKQDVCKTCTROPFSLAREACRSYSHVVLKEVRFF 180
Db 120 LFQWAVSRALCSPACSAFPRGAIKQDVCKTCTCARQPFSLAREACRSYSHVVLKEVRFF 179
Qy 181 NLQVLYPELLSDPALNLRIVHLVRDPRAVLRSREAAGPILARDNGIVLTGNGKWEADPHL 240
Db 180 NLQVLYPELLSDPALNLRIVHLVRDPRAVLRSREAAGPILARDNGIVLTGNGKWEADPHL 239
Qy 241 RLIREVCRSHVRIAEEATLKP-PPFLRGYRLVRPFEDLAREPLAEIRALYAFTGLTLPOL 300
Db 240 RLIREVCRSHVRIAEEATLKP-PPFLRGYRLVRPFEDLAREPLAEIRALYAFTGLTLPOL 299
Qy 301 EAWIHNITHGSG 312
Db 300 EAWIHNITHGSG 311

RESULT 12
US-09-816-825-2
; Sequence 2, Application US/09816825
; Patent No. US20010051370A1
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CON
; CURRENT APPLICATION NUMBER: US/09/816,825
; CURRENT FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 09/045,284
; PRIOR FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-816-825-2

Query Match 49.1%; Score 1008; DB 9; Length 386;
Best Local Similarity 52.3%; Pred. No. 9.1e-96;
Matches 206; Conservative 56; Mismatches 104; Indels 28; Gaps 7;

Qy 14 LLLAQTCLLLFIISRP-----GPSSPAGGEDRVHVLVLSWRSGSSFLGQ 59
Db 1 MLLPKQKLLFLVLSQMAILALFFHMYSHNISLSMKAQPERMHVVLVLSWRSGSSFLVGQ 60
Qy 60 LFSQHPDVFYLMPEAWHVTTLTSGSAATLHMAVRDLMSIFLDCMDVDFDAYM-POSRL 118
Db 61 LFGQHPDVFYLMPEAWHVTTLTSGSAATLHMAVRDLMSIFLDCMDVDFDAYMPCPRRQ 120
Qy 119 SAFPNWATSRALCSPACSAFPRGTISKQDVCKTCTROPFSLAREACRSYSHVVLKEVR 178
Db 121 SSLFQWENSRALCSPACSAFPRGTISKQDVCKTCTROPFSLAREACRSYSHVVLKEVR 180
Qy 179 FENLQVLYPELLSDPALNLRIVHLVRDPRAVLRSREAAGPILARDNGIVLTGN-GKWVEAD 237
Db 181 FENLQVLYPELLSDPALNLRIVHLVRDPRAVLRSREAAGPILARDNGIVLTGN-GKWVEAD 240
Qy 238 PHRLIREVCRSHVRIAEEATLKP-PPFLRGYRLVRPFEDLAREPLAEIRALYAFTGLTL 296
Db 241 QPYVMQVICSQLEIYK--TIQSLPKALQERYLLVRYEDLAPAPVAQTSRMVFEVGLF 298
Qy 297 TPOLAWIHNITHGSGIGKPIEAFHTSSRNARVNSQAWRHALPFTKILRVQEVACAGALQL 356
Db 299 LPHLQVWVNIIRGKMGD--HAFHTNARDALNVSOAWRWSLPEYKVSRLQKACGDMNL 356
Qy 357 LGVRPVYSADQORDLTLDLVLPRGPDHFSWASPD 390
Db 357 LGVRHVRSEQRNLLDLL-----STWTVPE 383

RESULT 13
US-10-007-262-1
; Sequence 1, Application US/10007262
; Publication No. US20020164748A1
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Tangemann, Kirsten
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CIP
; CURRENT APPLICATION NUMBER: US/10/007,262
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/190,911
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-12
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 386
; TYPE: PRT
; ORGANISM: H. sapiens
US-10-007-262-1

Query Match 49.1%; Score 1008; DB 13; Length 386;
Best Local Similarity 52.3%; Pred. No. 9.1e-96;
Matches 206; Conservative 56; Mismatches 104; Indels 28; Gaps 7;

Qy 14 LLLAQTCLLLFIISRP-----GPSSPAGGEDRVHVLVLSWRSGSSFLGQ 59
Db 1 MLLPKQKLLFLVLSQMAILALFFHMYSHNISLSMKAQPERMHVVLVLSWRSGSSFLVGQ 60
Qy 60 LFSQHPDVFYLMPEAWHVTTLTSGSAATLHMAVRDLMSIFLDCMDVDFDAYM-POSRL 118
Db 61 LFGQHPDVFYLMPEAWHVTTLTSGSAATLHMAVRDLMSIFLDCMDVDFDAYMPCPRRQ 120
Qy 119 SAFPNWATSRALCSPACSAFPRGTISKQDVCKTCTROPFSLAREACRSYSHVVLKEVR 178
Db 121 SSLFQWENSRALCSPACSAFPRGTISKQDVCKTCTROPFSLAREACRSYSHVVLKEVR 180
Qy 179 FENLQVLYPELLSDPALNLRIVHLVRDPRAVLRSREAAGPILARDNGIVLTGN-GKWVEAD 237
Db 181 FENLQVLYPELLSDPALNLRIVHLVRDPRAVLRSREAAGPILARDNGIVLTGN-GKWVEAD 240
Qy 238 PHRLIREVCRSHVRIAEEATLKP-PPFLRGYRLVRPFEDLAREPLAEIRALYAFTGLTL 296
Db 241 QPYVMQVICSQLEIYK--TIQSLPKALQERYLLVRYEDLAPAPVAQTSRMVFEVGLF 298
Qy 297 TPOLAWIHNITHGSGIGKPIEAFHTSSRNARVNSQAWRHALPFTKILRVQEVACAGALQL 356
Db 299 LPHLQVWVNIIRGKMGD--HAFHTNARDALNVSOAWRWSLPEYKVSRLQKACGDMNL 356
Qy 357 LGVRPVYSADQORDLTLDLVLPRGPDHFSWASPD 390
Db 357 LGVRHVRSEQRNLLDLL-----STWTVPE 383

RESULT 14
US-10-841-707-6
; Sequence 6, Application US/10841707
; Publication No. US20040202649A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Minoru
```

```

; APPLICANT: Yeh, Jiunn-Chern
; APPLICANT: Hiraoka, Nobuyoshi
; TITLE OF INVENTION: Identification of the Meca-79 Antigen
; TITLE OF INVENTION: and Related Methods of Treating L-Selectin-Mediated
; TITLE OF INVENTION: Conditions
; FILE REFERENCE: P-LJ 4149
; CURRENT APPLICATION NUMBER: US/10/841,707
; CURRENT FILING DATE: 2004-05-06
; PRIOR APPLICATION NUMBER: US/09/569,320A
; PRIOR FILING DATE: 2000-05-11
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-841-707-6

Query Match      48.9%; Score 1003.5; DB 16; Length 380;
Best Local Similarity 53.1%; Pred. No. 2.6e-95;
Matches 206; Conservative 55; Mismatches 102; Indels 25; Gaps 8;

Qy 13 VLLAQTCLLLF-----IISPGSSPAGGEDRVHVLVLSWRSGSSFLGQLFSQHP 65
Db 5 LFLVSQMAILALFFHMYSHNISLSMKAQP-----ERMHVLVLSWRSGSSFLGQLFSQHP 60
Qy 66 DVFLMEPAHWMTTLSQGSAAATLHMAVRDLMSIFLCMDVDFDAYM-PQSRNLSAFENW 124
Db 61 DVFLMEPAHWMTTQKSTAWMLHMAVRDLIRAVFLCDMSVFDAYMEPGPRROSSLQW 120
Qy 125 ATRALCSPPACSAFPRGTISKQDVCKTLCTRPFPFLAREACRSYSHVVLKEVFFNLQV 184
Db 121 ENSRALCSAPACDIIPQDEIIPRAHCRLLCSQFPVEVVEKACRSYSHVVLKEVFFNLQ 180
Qy 185 LYPILSDPALNLRVHVRDPAVLRSREAGPILARDNGIVLGTN-GKWVEADPHRLI 243
Db 181 LYPILSDPALNLRVHVRDPAVLRSREAGPILARDNGIVLGTN-GKWVEADPHRLI 240
Qy 244 REVCRSHVRIAEATLKP-PPLFGRVRLVRFEDLAREPLAEIRALYAFTGLTLPOL 302
Db 241 QVICOSQLEIYK--TIQSLPKALQERYLLVRYEDLARAPVAQTSRMVFEVGLFPLHQT 298
Qy 303 WIHNTIGSGIGKPIEAFHTSSRNARNVSQAWRHALPFTKILRVQEVCAQALQLLYRPV 362
Db 299 VWNITRKGKMGD--HAFHTNARDALNVSQAWRWSLPYKVSRLQKACGDAMNLLGYRHV 356
Qy 363 YSADQORDLTLDLVLPRGPDHFSWASPD 390
Db 357 RSEQEQRNLLDLL-----STWTVPE 377

```

```

RESULT 15
US-10-427-631-11
; Sequence 11, Application US/10427631
; Publication No. US20030175923A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION; TANG, Y. TOM;
; APPLICANT: CORLEY, Neil C.; GUEGLER, Karl J.;
; APPLICANT: BAUGHN, Maria R.; LAL, Preeti G.;
; APPLICANT: YUE, Henry; HILLMAN, Jennifer L.;
; APPLICANT: AZIMZAI, Yalda
; TITLE OF INVENTION: HUMAN TRANSFERASE PROTEINS
; FILE REFERENCE: PP-0592-1 DIV
; CURRENT APPLICATION NUMBER: US/10/427,631
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 09/786,240
; PRIOR FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: PCT/US99/20989
; PRIOR FILING DATE: 1999-09-09
; PRIOR APPLICATION NUMBER: US 60/172,220
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: US 60/155,248
; PRIOR FILING DATE: 1998-11-04

```

```

; PRIOR APPLICATION NUMBER: US 60/133,642
; PRIOR FILING DATE: 1999-05-11
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PERL Program
; SEQ ID NO 11
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030175923A1 2617407CD1
US-10-427-631-11

Query Match      47.3%; Score 970; DB 14; Length 386;
Best Local Similarity 51.3%; Pred. No. 8.3e-92;
Matches 202; Conservative 52; Mismatches 112; Indels 28; Gaps 7;

Qy 14 LLLAQTCLLLFIISRP-----GSSSPAGGEDRVHVLVLSWRSGSSFLGQ 59
Db 1 MLLPKMKKLLFLVVSQMAILALFFHMYSHNISLSMKAQPERMHVLVLSWRSGSSFLVGQ 60
Qy 60 LFSQHPDVFYLMPEPAHWMTTLSQGSAAATLHMAVRDLMSIFLCMDVDFDAYM-PQSRNL 118
Db 61 LFSQHPDVFYLMPEPAHWMTTQKSTAWMLHMAVRDLIRAVFLCDMSVFDAYMEPGPRQ 120
Qy 119 SAFPNWATSRALCSPPACSAFPRGTISKQDVCKTLCTRPFPFLAREACRSYSHVVLKEVR 178
Db 121 SSLFQWENSRALCSAPACDIIPQDESSPGLTAGSCAVNSPLKLEKACRSYSHVVLKEVR 180
Qy 179 FPNLOVLYPLSDPALNLRVHVRDPAVLRSREAGPILARDNGIVLGTN-GKWVEAD 237
Db 181 FPNLOVLYPLSDPALNLRVHVRDPAVLRSREAGPILARDNGIVLGTN-GKWVEAD 240
Qy 238 PHLRLEIVCRSHVRIAEATLKP-PPLFGRVRLVRFEDLAREPLAEIRALYAFTGLTL 296
Db 241 QPYVMQVICSQLEIYK--TIQSLPKALQERYLLVRYEDLARAPVAQTSRMVFEVGLF 298
Qy 297 TPQLEAWIHNTIGSGIGKPIEAFHTSSRNARNVSQAWRHALPFTKILRVQEVCAQALQL 356
Db 299 LPHLQTVWNITRKGKMGD--HAFHTNARDALNVSQAWRWSLPYKVSRLQKACGDAMNL 356
Qy 357 LGTRPVYSADQORDLTLDLVLPRGPDHFSWASPD 390
Db 357 LGYRHRVSRSEQEQRNLLDLL-----STWTVPE 383

```

Search completed: June 23, 2005, 09:27:53
Job time : 50.3029 secs

THIS PAGE BLANK (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2005, 08:34:49 ; Search time 14.6998 Seconds
(without alignments)
1980.512 Million cell updates/sec

Title: US-10-697-828-8

Perfect score: 2051

Sequence: 1 MWLPFRSSKTVTVLLAQTT.....LTLDLVLPRGPDHFSWASPD 390

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:**

1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*

2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*

3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*

4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*

5: /cgn2_6/ptodata/1/iaa/FCRUS_COMB.pep.*

6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|--|
| 1 | 2051 | 100.0 | 390 | 4 | US-09-949-016-6813 Sequence 6813, Ap |
| 2 | 2051 | 100.0 | 431 | 4 | US-09-949-016-6893 Sequence 8893, Ap |
| 3 | 1729.5 | 84.3 | 395 | 4 | US-09-949-016-7011 Sequence 7011, Ap |
| 4 | 1008 | 49.1 | 386 | 3 | US-09-045-284A-2 Sequence 2, Appli |
| 5 | 1008 | 49.1 | 386 | 3 | US-09-190-911-1 Sequence 1, Appli |
| 6 | 970 | 47.3 | 386 | 4 | US-09-786-240-11 Sequence 11, Appl |
| 7 | 665 | 32.4 | 483 | 3 | US-09-263-023-2 Sequence 2, Appli |
| 8 | 665 | 32.4 | 483 | 4 | US-09-471-867-2 Sequence 2, Appli |
| 9 | 659.5 | 32.2 | 608 | 4 | US-09-949-016-9449 Sequence 9449, Ap |
| 10 | 656.5 | 32.0 | 484 | 3 | US-09-263-023-4 Sequence 4, Appli |
| 11 | 656.5 | 32.0 | 484 | 4 | US-09-471-867-4 Sequence 4, Appli |
| 12 | 656.5 | 32.0 | 531 | 4 | US-09-949-016-6471 Sequence 6471, Ap |
| 13 | 577.5 | 28.2 | 479 | 4 | US-08-899-514-2 Sequence 2, Appli |
| 14 | 540.5 | 26.4 | 458 | 2 | US-08-655-878-2 Sequence 2, Appli |
| 15 | 513.5 | 25.0 | 411 | 3 | US-09-015-188-2 Sequence 2, Appli |
| 16 | 105 | 5.1 | 359 | 3 | US-09-150-133-11 Sequence 11, Appl |
| 17 | 105 | 5.1 | 359 | 3 | US-09-150-141-11 Sequence 11, Appl |
| 18 | 105 | 5.1 | 359 | 3 | US-09-374-493-11 Sequence 11, Appl |
| 19 | 105 | 5.1 | 359 | 3 | US-09-374-824-11 Sequence 11, Appl |
| 20 | 105 | 5.1 | 359 | 3 | US-09-374-492-11 Sequence 11, Appl |
| 21 | 105 | 5.1 | 359 | 4 | US-09-785-343-11 Sequence 11, Appl |
| 22 | 105 | 5.1 | 359 | 4 | US-10-411-976-11 Sequence 11, Appl |
| 23 | 97 | 4.7 | 199 | 4 | US-09-252-901A-28497 Sequence 28497, A |
| 24 | 91.5 | 4.5 | 242 | 4 | US-09-248-796A-20943 Sequence 20943, A |
| 25 | 90.5 | 4.4 | 380 | 3 | US-09-150-133-9 Sequence 9, Appli |
| 26 | 90.5 | 4.4 | 380 | 3 | US-09-150-141-9 Sequence 9, Appli |
| 27 | 90.5 | 4.4 | 380 | 3 | US-09-374-493-9 Sequence 9, Appli |

ALIGNMENTS

RESULT 1

US-09-949-016-6813

; Sequence 6813, Application US/09949016

; Patent No. 6812339

; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

; FILE REFERENCE: CLO01307

; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 2000-04-14

; PRIOR APPLICATION NUMBER: 60/241,755

; PRIOR FILING DATE: 2000-10-20

; PRIOR APPLICATION NUMBER: 60/237,768

; PRIOR FILING DATE: 2000-10-03

; PRIOR APPLICATION NUMBER: 60/231,498

; PRIOR FILING DATE: 2000-09-08

; NUMBER OF SEQ ID NOS: 207012

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 6813

; LENGTH: 390

; TYPE: PRT

; ORGANISM: Human

US-09-949-016-6813

Query Match 100.0%; Score 2051; DB 4; Length 390;

Best Local Similarity 100.0%; Pred. No. 2.5e-217;

Matches 390; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MWLPFRSSKTVTVLLAQTTCLLLFIISRPSPSPAGEDRVHVLVLSWSSGSSFLGQL 60

Db 1 MWLPFRSSKTVTVLLAQTTCLLLFIISRPSPSPAGEDRVHVLVLSWSSGSSFLGQL 60

Qy 61 FSOHPDVFYLMPEAHVWTTLSQGSAAATLHMAVRDLMRSIFCLDMDFDAYMPSQSNLSA 120

Db 61 FSOHPDVFYLMPEAHVWTTLSQGSAAATLHMAVRDLMRSIFCLDMDFDAYMPSQSNLSA 120

Qy 121 FFWNATSRALCSPACSAFFRGITISKQDVCKTLCTROPFSLAREACRSYSHVVLKEVRFF 180

Db 121 FFWNATSRALCSPACSAFFRGITISKQDVCKTLCTROPFSLAREACRSYSHVVLKEVRFF 180

Qy 181 NLQVLYPLLSDDPALNLRIVHLVRDPRRAVLRSSREAAGPILARDNGIVLGTNGKWEADPHL 240

Db 181 NLQVLYPLLSDDPALNLRIVHLVRDPRRAVLRSSREAAGPILARDNGIVLGTNGKWEADPHL 240

Qy 241 RLIREVCRSHVRIAETAATLKPPLRCGRYLRVRFEDLAREPLAERIALYAFYGLTTLTPQL 300

Db 241 RLIREVCRSHVRIAETAATLKPPLRCGRYLRVRFEDLAREPLAERIALYAFYGLTTLTPQL 300

Qy 301 EAWTHNTHGSGIGKPIEAFHTSSRNARNVSQAWRHALPFTTKILRVQEVCAQALQLLGYR 360

Db 301 EAWTHNTHGSGIGKPIEAFHTSSRNARNVSQAWRHALPFTTKILRVQEVCAQALQLLGYR 360

| | | | | | | |
|----|------|-----|-----|---|----------------------|-------------------|
| 28 | 90.5 | 4.4 | 380 | 3 | US-09-374-824-9 | Sequence 9, Appli |
| 29 | 90.5 | 4.4 | 380 | 3 | US-09-374-492-9 | Sequence 9, Appli |
| 30 | 90.5 | 4.4 | 380 | 4 | US-09-785-343-9 | Sequence 9, Appli |
| 31 | 90.5 | 4.4 | 380 | 4 | US-10-411-976-9 | Sequence 9, Appli |
| 32 | 90.5 | 4.4 | 984 | 2 | US-08-673-789-9 | Sequence 9, Appli |
| 33 | 90 | 4.4 | 575 | 4 | US-09-252-991A-32986 | Sequence 32986, A |
| 34 | 88 | 4.3 | 288 | 4 | US-09-818-780-17 | Sequence 17, Appl |
| 35 | 88 | 4.3 | 268 | 4 | US-09-818-780-94 | Sequence 94, Appl |
| 36 | 87 | 4.2 | 255 | 4 | US-09-902-540-15692 | Sequence 15692, A |
| 37 | 86.5 | 4.2 | 376 | 3 | US-09-150-133-7 | Sequence 7, Appli |
| 38 | 86.5 | 4.2 | 376 | 3 | US-09-150-141-7 | Sequence 7, Appli |
| 39 | 86.5 | 4.2 | 376 | 3 | US-09-374-493-7 | Sequence 7, Appli |
| 40 | 86.5 | 4.2 | 376 | 3 | US-09-374-824-7 | Sequence 7, Appli |
| 41 | 86.5 | 4.2 | 376 | 3 | US-09-374-492-7 | Sequence 7, Appli |
| 42 | 86.5 | 4.2 | 376 | 4 | US-09-785-343-7 | Sequence 7, Appli |
| 43 | 86.5 | 4.2 | 376 | 4 | US-10-411-976-7 | Sequence 7, Appli |
| 44 | 86.5 | 4.2 | 377 | 3 | US-09-150-133-5 | Sequence 5, Appli |
| 45 | 86.5 | 4.2 | 377 | 3 | US-09-150-141-5 | Sequence 5, Appli |

```

; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7011
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Human
; ORGANISM: Homo sapiens
US-09-949-016-8893

Query Match      84.3%; Score 1729.5; DB 4; Length 395;
Best Local Similarity 85.8%; Pred. No. 7.4e-182;
Matches 333; Conservative 17; Mismatches 37; Indels 1; Gaps 1;

Db      301 EAMHNTHSGGIGKPIEAFTSSRNARNVQAWRHALPTKILRVOECAGALQLLYR 360
Qy      361 PVSADQOQDLTLDLVLPRGPDHFSWASPD 390
Db      361 PVSADQOQDLTLDLVLPRGPDHFSWASPD 390

RESULT 2
US-09-949-016-8893
; Sequence 8893, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8893
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Human
; ORGANISM: Homo sapiens
US-09-949-016-8893

Query Match      100.0%; Score 2051; DB 4; Length 431;
Best Local Similarity 100.0%; Pred. No. 2.9e-217;
Matches 390; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      301 EAMHNTHSGGIGKPIEAFTSSRNARNVQAWRHALPTKILRVOECAGALQLLYR 360
Qy      361 PVSADQOQDLTLDLVLPRGPDHFSWASPD 390
Db      361 PVSADQOQDLTLDLVLPRGPDHFSWASPD 390

RESULT 3
US-09-949-016-7011
; Sequence 7011, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

```

```

; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7011
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Human
; ORGANISM: Homo sapiens
US-09-949-016-7011

Query Match      84.3%; Score 1729.5; DB 4; Length 395;
Best Local Similarity 85.8%; Pred. No. 7.4e-182;
Matches 333; Conservative 17; Mismatches 37; Indels 1; Gaps 1;

Qy      1 MMLPRFSSKTVTVLLAQTTCLLLFIIISRPSPAGGEDRVHVLVLSWRSGSFLGQL 60
Db      1 MMLPRVSSSTAVTALLAQ-TFLLLFLVSRPGSPSPAGGEARVHVLVLSWRSGSFLGQL 59
Qy      61 FSQHPDVFYLMPEPAWHVWTTLSQSSAATLHMAVRDLMSIFLCMDVDFDAYMPSQSNLSA 120
Db      60 FNQHPDVFYLMPEPAWHVWTTLSQSSAATLHMAVRDLVRSVFLCDMDVDFDAYLPPWRNLS 119
Qy      121 FFWNATSRALCSPPCSAFPGRGTISKQDVCKTLCTROPFSLAREACRSYSHVVVKVRF 180
Db      120 LQWAVSRALCSPPCSAFPGRGAISSSEAVCKPLCAQSFILAREACRSYSHVVVKVRF 179
Qy      181 NLQVLYPLSDPALNLRIVHLVRDPRAVLRSREAAAGPILARDNGIVLGTNGKWEADPHL 240
Db      180 NLQVLYPLSDPALNLRIVHLVRDPRAVLRSREQTAKALARDNGIVLGTNGTWVEADPGL 239
Qy      241 RLIREVCRSHVRIAEAAATLKPPPLRGYRLVRFEFLAREPLAEIRALYAFTGLTLPQL 300
Db      240 RVREVCRRSHVRIAEAAATLKPPPLRGYRLVRFEFLAREPLAEIRALYAFTGLTLPQL 299
Qy      301 EAMHNTHSGGIGKPIEAFTSSRNARNVQAWRHALPTKILRVOECAGALQLLYR 360
Db      300 EAMHNTHSGGIGKPIEAFTSSRNARNVQAWRHALPTKILRVOECAGALQLLYR 359
Qy      361 PVSADQOQDLTLDLVLPRGPDHFSWAS 388
Db      360 PVSDEQQRNALDLVLPRLNGFTWAS 387

RESULT 4
US-09-045-284A-2
; Sequence 2, Application US/09045284A
; Patent No. 6265192
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107US1
; CURRENT APPLICATION NUMBER: US/09/045,284A
; CURRENT FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
; ORGANISM: Homo sapiens
US-09-045-284A-2

Query Match      49.1%; Score 1008; DB 3; Length 386;
Best Local Similarity 52.3%; Pred. No. 2.7e-102;

```

Matches 206; Conservative 56; Mismatches 104; Indels 28; Gaps 7;

Qy 14 LLLAQTCLLLFIISRP-----GPSSPAGGEDRVHVLVLSWRSGSFLGQ 59
 Db 1 MLLPKMKLLFLVSQMAILALPFHMYSHNITSSLSMKAQPERMHVVLSSWRSGSFLGQ 60
 Qy 60 LFSQHPDVFYLMPEPAWHVMTTSLQSGSAATHMAVRDLMSIFLCMDVFDAYM-PQSRNL 118
 Db 61 LFGQHPDVFYLMPEPAWHVMTTSLQSGSAATHMAVRDLMSIFLCMDVFDAYM-PQSRNL 120
 Qy 119 SAFFNWATSRALCPSPACSAFPRGTISKQDVCKLTCTROPESLAREACRSYSHVVLEKEVR 178
 Db 121 SSLFQWNSRSLCSPACDIIPQDEIIIPRAHCRLLCSQPPFVVEKACRSYSHVVLEKEVR 180
 Qy 179 FFNQLVLYPLSDPALNLRIVHLVRDPRVLRSEAAAGPILARDNGIVLGTN-GKWVEAD 237
 Db 181 FFNQLVLYPLSDPALNLRIVHLVRDPRVLRSEAAAGPILARDNGIVLGTN-GKWVEAD 240
 Qy 238 PHRLIREVCRSHVRIAAATLKP-PPFLRGRYLVRFEDLAREPLAEIRALYAFTGLTL 296
 Db 241 QPYVVMQVICSQLEIVK--TIQSLPKALQERYLLVRYEDLARAPVAQTSRMVBFVGLGF 298
 Qy 297 TPQLEAMHNITHSGGIGKPIEAFHTSSRNARVNSQAWRHLPFTKILRVQEVCGAGLQL 356
 Db 299 LPHLOTWVHNITRKGMD--HAFHTNARDALNVSOAWRWSLPYKVSRLQKACGDAMNL 356
 Qy 357 LGYRPVVSADQORDLTLDLVLPRGPDHFWASPD 390
 Db 357 LGYRHRVSEQQRNLLDLL-----STWTVE 383

RESULT 5

US-09-190-911-1

; Sequence 1, Application US/09190911
 ; Patent No. 6365365
 ; GENERAL INFORMATION:
 ; APPLICANT: Bistrup, Annette
 ; APPLICANT: Rosen, Steven D.
 ; APPLICANT: Tangemann, Kirsten
 ; APPLICANT: Hemmerich, Stefan
 ; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
 ; FILE REFERENCE: 6510-107CIP
 ; CURRENT APPLICATION NUMBER: US/09/190,911
 ; EARLIER FILING DATE: 1998-11-12
 ; EARLIER FILING DATE: 1998-03-20
 ; NUMBER OF SEQ ID NOS: 8
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 1
 ; LENGTH: 386
 ; TYPE: PRT
 ; ORGANISM: H. sapiens
 US-09-190-911-1

Query Match 49.1%; Score 1008; DB 3; Length 386;
 Best Local Similarity 52.3%; Pred. No. 2.7e-102;
 Matches 206; Conservative 56; Mismatches 104; Indels 28; Gaps 7;

Qy 14 LLLAQTCLLLFIISRP-----GPSSPAGGEDRVHVLVLSWRSGSFLGQ 59
 Db 1 MLLPKMKLLFLVSQMAILALPFHMYSHNITSSLSMKAQPERMHVVLSSWRSGSFLGQ 60
 Qy 60 LFSQHPDVFYLMPEPAWHVMTTSLQSGSAATHMAVRDLMSIFLCMDVFDAYM-PQSRNL 118
 Db 61 LFGQHPDVFYLMPEPAWHVMTTSLQSGSAATHMAVRDLMSIFLCMDVFDAYM-PQSRNL 120
 Qy 119 SAFFNWATSRALCPSPACSAFPRGTISKQDVCKLTCTROPESLAREACRSYSHVVLEKEVR 178
 Db 121 SSLFQWNSRSLCSPACDIIPQDEIIIPRAHCRLLCSQPPFVVEKACRSYSHVVLEKEVR 180
 Qy 179 FFNQLVLYPLSDPALNLRIVHLVRDPRVLRSEAAAGPILARDNGIVLGTN-GKWVEAD 237
 Db 181 FFNQLVLYPLSDPALNLRIVHLVRDPRVLRSEAAAGPILARDNGIVLGTN-GKWVEAD 240

Qy 238 PHRLIREVCRSHVRIAAATLKP-PPFLRGRYLVRFEDLAREPLAEIRALYAFTGLTL 296
 Db 241 QPYVVMQVICSQLEIVK--TIQSLPKALQERYLLVRYEDLARAPVAQTSRMVBFVGLGF 298
 Qy 297 TPQLEAMHNITHSGGIGKPIEAFHTSSRNARVNSQAWRHLPFTKILRVQEVCGAGLQL 356
 Db 299 LPHLOTWVHNITRKGMD--HAFHTNARDALNVSOAWRWSLPYKVSRLQKACGDAMNL 356
 Qy 357 LGYRPVVSADQORDLTLDLVLPRGPDHFWASPD 390
 Db 357 LGYRHRVSEQQRNLLDLL-----STWTVE 383

RESULT 6

US-09-786-240-11
 ; Sequence 11, Application US/09786240
 ; Patent No. 6558935
 ; GENERAL INFORMATION:
 ; APPLICANT: INCYTE PHARMACEUTICALS, INC.
 ; APPLICANT: TANG, Y. Tom
 ; APPLICANT: CORLEY, Neil C.
 ; APPLICANT: GUEGLER, Karl J.
 ; APPLICANT: BAUGHN, Mariah R.
 ; APPLICANT: LAL, Preeti
 ; APPLICANT: YUE, Henry
 ; APPLICANT: HILLMAN, Jennifer L.
 ; APPLICANT: AZIMZAI, Valda
 ; TITLE OF INVENTION: HUMAN TRANSFERASE PROTEINS
 ; FILE REFERENCE: PF-0592 PCT
 ; CURRENT APPLICATION NUMBER: US/09/786,240
 ; CURRENT FILING DATE: 2002-03-12
 ; PRIOR APPLICATION NUMBER: 09/150,657; unassigned; 09/186,779; unassigned; 60/133,642
 ; PRIOR FILING DATE: 1998-09-10; 1998-09-10; 1998-11-04; 1998-11-04; 1999-05-11
 ; NUMBER OF SEQ ID NOS: 33
 ; SOFTWARE: PERL Program
 ; SEQ ID NO 11
 ; LENGTH: 386
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; OTHER INFORMATION: Incyte ID No. 6558935 2617407CD1
 US-09-786-240-11

Query Match 47.3%; Score 970; DB 4; Length 386;
 Best Local Similarity 51.3%; Pred. No. 4.3e-98;
 Matches 202; Conservative 52; Mismatches 112; Indels 28; Gaps 7;

Qy 14 LLLAQTCLLLFIISRP-----GPSSPAGGEDRVHVLVLSWRSGSFLGQ 59
 Db 1 MLLPKMKLLFLVSQMAILALPFHMYSHNITSSLSMKAQPERMHVVLSSWRSGSFLGQ 60
 Qy 60 LFSQHPDVFYLMPEPAWHVMTTSLQSGSAATHMAVRDLMSIFLCMDVFDAYM-PQSRNL 118
 Db 61 LFGQHPDVFYLMPEPAWHVMTTSLQSGSAATHMAVRDLMSIFLCMDVFDAYM-PQSRNL 120
 Qy 119 SAFFNWATSRALCPSPACSAFPRGTISKQDVCKLTCTROPESLAREACRSYSHVVLEKEVR 178
 Db 121 SSLFQWNSRSLCSPACDIIPQDEIIIPQDESPGLTAGCAVNSPLKLEKACRSYSHVVLEKEVR 180
 Qy 179 FFNQLVLYPLSDPALNLRIVHLVRDPRVLRSEAAAGPILARDNGIVLGTN-GKWVEAD 237
 Db 181 FFNQLVLYPLSDPALNLRIVHLVRDPRVLRSEAAAGPILARDNGIVLGTN-GKWVEAD 240
 Qy 238 PHRLIREVCRSHVRIAAATLKP-PPFLRGRYLVRFEDLAREPLAEIRALYAFTGLTL 296
 Db 241 QPYVVMQVICSQLEIVK--TIQSLPKALQERYLLVRYEDLARAPVAQTSRMVBFVGLGF 298
 Qy 297 TPQLEAMHNITHSGGIGKPIEAFHTSSRNARVNSQAWRHLPFTKILRVQEVCGAGLQL 356
 Db 299 LPHLOTWVHNITRKGMD--HAFHTNARDALNVSOAWRWSLPYKVSRLQKACGDAMNL 356


```

; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 9449
; LENGTH: 608
; TYPE: PRT
; ORGANISM: Human
; US-09-949-016-9449

Query Match          32.2%; Score 659.5; DB 4; Length 608;
Best Local Similarity 39.1%; Pred. No. 1.5e-63;
Matches 151; Conservative

Qy 31 GPSPAG-----CEDRVHVLVSSWSSGSGFLQSFQHPDVFYLMPEPAWHVWTTLL 81
Db 223 GVAAPPNGTRGTGGVGDQRQLVVFVTTWSSGSGFFGELFNQNPVEVFLYEPVHVWVQKL 282

Qy 82 SQSAATLHMVRLMRSIFLCMDVDFDAYMPQ---SRNLS--AFFNWATSRALCSPAC 136
Db 283 YPGDAVSLOGAARDMLSALYRCDSLVSFQLYSPAGSGGRNLTTLGIFGAATNKVVCSSPLC 342

Qy 137 SAPPRGTISKQD--VCKTCLTCTROPFSLAREACRSYSHVVLKEVRFNQLVLYPLLSDPAL 194
Db 343 PAYRKEVVGVLVDVCKK-CPQELARPEBECKRYRTLVIKGVRFVDFVAVLAPLRLDPAL 401

Qy 195 NLRIVHLVRDPRVLR-----REAGPILARD-----NGIVLTNGKWV 234
Db 402 DLKVIHLVRDPRVAVSSIRSRHGLIRESLQVRSRDPRAHRMPFLEAGHKLGAKEGV 461

Qy 235 --EADPH-LRLREVCHSHVRIAEAAATLKPPFLGRYLRVRFEDLAREPLAEIRALYAF 291
Db 462 GGPADYHALGAMEVICNSMAKTLQTA-LQPPDWLQGHYLVVRYEDLVGDPVKTLRRVYDF 520

Qy 292 TGLTLTLPQLEAHINITHGSG-IGKPIEAFTSSRNARNVSOAQRHALPFTKILRVOEVC 350
Db 521 VGLLVSPMEQFALNMTSGSSSKP---FVVSARNATQAAANWRTALTFOQIKQVEEFC 577

Qy 351 AGALQLLGYRPVVSADQORDITLDLV 376
Db 578 YQPMVAVLGYERVNSPEEVKLSKTL 603

RESULT 10
US-09-263-023-4
; Sequence 4, Application US/09263023
; Patent No. 6037159
; GENERAL INFORMATION:
; APPLICANT: Uchimura, Kenji
; APPLICANT: Muramatsu, Hideki
; APPLICANT: Kadomatsu, Kenji
; APPLICANT: Kannagi, Reiji
; APPLICANT: Habuchi, Osami
; APPLICANT: Muramatsu, Takashi
; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-SULFOTRANSFERASE AND
; FILE REFERENCE: TOYAMA1.001AUS
; CURRENT APPLICATION NUMBER: US/09/263,023
; EARLIER FILING DATE: 1999-03-05
; EARLIER FILING DATE: 1998-03-05
; EARLIER FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-263-023-4

Query Match          32.0%; Score 656.5; DB 3; Length 484;
Best Local Similarity 39.7%; Pred. No. 2.3e-63;
Matches 149; Conservative

Qy 34 SPAGGEDRVH-VLVLSWSSGSGFLQSFQHPDVFYLMPEPAWHVWTTLLSQGSAATLHMA 92
Db 110 APEGVGDKRHHMYVFTTWSSGSGFFGELFNQNPVEVFLYEPVHVWVQKLYPGDAVSLOGA 169

Qy 93 VRDLMSIFLCMDVDFDAYMPQ---SRNLS--AFFNWATSRALCSPACSAFPRGTISKQ 147
Db 170 ARDMLSALYRCDSLVSFQLYSPAGSGGRNLTTLGIFGAATNKVVCSSPLCPAYRKEVVGLV 229

Qy 148 D--VCKTCLTCTROPFSLAREACRSYSHVVLKEVRFNQLVLYPLLSDPALNLRIVHLVRDP 205
Db 230 DDRVCKK-CPQELARPEBECKRYRTLVIKGVRFVDFVAVLAPLRLDPALDVKIHLVRDP 288

Qy 206 RAVLRS-----REAGPILARD-----NGIVLTNGKWV--EADPH-LRL 242
Db 289 RAVASSIRSRHGLIRESLQVRSRDPRAHRMPFLEAGHKLGAKEGVGPGADYHALGA 348

Qy 243 IREVCVSHVRIAEAAATLKPPFLGRYLRVRFEDLAREPLAEIRALYAFVLTLPQLEA 302
Db 349 MEVICNSMAKTLQTA-LQPPDWLQGHYLVVRYEDLVGDPVKTLRRVYDFVGLLVSPMEQ 407

Qy 303 WIHNITHGSG-IGKPIEAFTSSRNARNVSOAQRHALPFTKILRVOEVCAGALQLLGYRP 361
Db 408 FALNMTSGSSSKP---FVVSARNATQAAANWRTALTFOQIKQVEEFCYQPMVAVLGYER 464

Qy 362 VYSADQORDITLDLV 376
Db 465 VNSPEEVKLSKTL 479

RESULT 11
US-09-471-867-4
; Sequence 4, Application US/09471867
; Patent No. 6455289
; GENERAL INFORMATION:
; APPLICANT: Uchimura, Kenji
; APPLICANT: Muramatsu, Hideki
; APPLICANT: Kadomatsu, Kenji
; APPLICANT: Kannagi, Reiji
; APPLICANT: Habuchi, Osami
; APPLICANT: Muramatsu, Takashi
; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-SULFOTRANSFERASE AND
; FILE REFERENCE: TOYAMA1.001AUS
; CURRENT APPLICATION NUMBER: US/09/471,867
; CURRENT FILING DATE: 1999-12-23
; EARLIER FILING DATE: 1999-03-05
; EARLIER FILING DATE: 1998-03-05
; EARLIER FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-471-867-4

Query Match          32.0%; Score 656.5; DB 4; Length 484;
Best Local Similarity 39.7%; Pred. No. 2.3e-63;
Matches 149; Conservative

Qy 34 SPAGGEDRVH-VLVLSWSSGSGFLQSFQHPDVFYLMPEPAWHVWTTLLSQGSAATLHMA 92
Db 110 APEGVGDKRHHMYVFTTWSSGSGFFGELFNQNPVEVFLYEPVHVWVQKLYPGDAVSLOGA 169

Qy 93 VRDLMSIFLCMDVDFDAYMPQ---SRNLS--AFFNWATSRALCSPACSAFPRGTISKQ 147
Db 170 ARDMLSALYRCDSLVSFQLYSPAGSGGRNLTTLGIFGAATNKVVCSSPLCPAYRKEVVGLV 229

Qy 148 D--VCKTCLTCTROPFSLAREACRSYSHVVLKEVRFNQLVLYPLLSDPALNLRIVHLVRDP 205

```

```

Db 230 DDRVCKK-CPQRLARFECECKRYTLVIGKVRVDFVAVLAPLRLDPALDLKVIHLVRDP 288
Qy 206 RAVLRS-----REAAGPILARD-----NGIVLTNGKWV--EADPH-LRL 242
Db 289 RAVASSIRSRHGILRESLQVRSRDRPRAHMPFLEAAGHKLGAKKEGVGGPADIHALGA 348
Qy 243 IREVCRRSHVRIAEATLKPPFLGRGRYLVRFEDLAREPLAEIRALYAFTGLTILTPOLEA 302
Db 349 MEVICNSMAKTLQTA-LQPPDWLQGHVLYVRYEDLVGDPVKTLRRVYDFVGLLVSPMEQ 407
Qy 303 WIHNITHGSG-IGKPIEAFHTSSRNARNVQAMRHALPFTKILRVQEVCGAGLQLLYGRP 361
Db 408 FALNMTSGSGSSSKP---FVVSARNATQANAMWRTALTFOQIKQVEBFCYQPMVAVLYGER 464
Qy 362 VYSADQQRDLTLDAV 376
Db 465 VNSPEEVKDLJSTLL 479

RESULT 12
US-09-949-016-6471
; Sequence 6471, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6471
; LENGTH: 531
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6471

Query Match 32.0%; Score 656.5; DB 4; Length 531;
Best Local Similarity 35.7%; Pred. No. 2.7e-63;
Matches 149; Conservative 68; Mismatches 121; Indels 37; Gaps 12;

Qy 34 SPAGEDRVH-VLSSWRSGSFLGQFSDHDPVFLMEPAWHVMTTLTSGGSAATLHMA 92
Db 157 APEGVGDKRHWYVFTTWRSGSFFGLFNQNPVEVFLYEPVHVWQKLYPGDAVSQGA 216
Qy 93 VRDLMSRIFCLDMDFDAYMPQ----SRNLS--AFFNWTARALCSPACSAFPRGTISKQ 147
Db 217 ARDMLSALRYCDLSVQLYSPAGSGGNLTTLGIFGAATNKVCSPLCPAYRKEVVGVL 276
Qy 148 D--VCKTLCTROPFSLAREACRSYSHVVLKEVRFENLQVLYPLSDPALNLRIVHLVRDP 205
Db 277 DDRVCKK-CPQRLARFECECKRYTLVIGKVRVDFVAVLAPLRLDPALDLKVIHLVRDP 335
Qy 206 RAVLRS-----REAAGPILARD-----NGIVLTNGKWV--EADPH-LRL 242
Db 336 RAVASSIRSRHGILRESLQVRSRDRPRAHMPFLEAAGHKLGAKKEGVGGPADIHALGA 395
Qy 243 IREVCRRSHVRIAEATLKPPFLGRGRYLVRFEDLAREPLAEIRALYAFTGLTILTPOLEA 302
Db 396 MEVICNSMAKTLQTA-LQPPDWLQGHVLYVRYEDLVGDPVKTLRRVYDFVGLLVSPMEQ 454
Qy 303 WIHNITHGSG-IGKPIEAFHTSSRNARNVQAMRHALPFTKILRVQEVCGAGLQLLYGRP 361
Db 455 FALNMTSGSGSSSKP---FVVSARNATQANAMWRTALTFOQIKQVEBFCYQPMVAVLYGER 511
Qy 362 VYSADQQRDLTLDAV 376

```

```

Db 512 VNSPEEVKDLSTLL 526

RESULT 13
US-08-899-514-2
; Sequence 2, Application US/08899514
; Patent No. 5910581
; GENERAL INFORMATION:
; APPLICANT: HASEUCHI, OSAMI
; APPLICANT: FUKUTA, MASAKAZU
; TITLE OF INVENTION: POLYPEPTIDE OF GLYCOSAMINOGLYCAN
; TITLE OF INVENTION: SULFOTRANSFERASE ORIGINATING FROM HUMAN AND DNA CODING
; TITLE OF INVENTION: FOR THE POLYPEPTIDE
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: KNOBBE, MARTENS, OLSON & BEAR, LLP
; STREET: 620 NEWPORT CENTER DRIVE, SIXTEENTH FLOOR
; CITY: NEWPORT BEACH
; STATE: CALIFORNIA
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/899,514
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: DANIEL E ALTMAN
; REGISTRATION NUMBER: 34,115
; REFERENCE/DOCKET NUMBER: TOYAM21.001AUS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 714 760 0404
; TELEFAX: 714 760 9502
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 479
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-899-514-2

Query Match 28.2%; Score 577.5; DB 2; Length 479;
Best Local Similarity 35.7%; Pred. No. 1.2e-54;
Matches 132; Conservative 65; Mismatches 134; Indels 39; Gaps 8;

Qy 32 PSSPAGGEDRVHVLSSWRSGSFLGQFSDHDPVFLMEPAWHVMTTLT-----QGSAA 87
Db 123 PPRPAVAGRRHVLMTATRTTSSFGVEFNQGNIFLFEPLWHIERTVSPSPGGANAA 182
Qy 88 TLHMAVRDLMSRIFCLDMDFDAY---MQSRNLSAFFNWTARALCSPACSAFPRGTI 144
Db 183 GSALVYRDVLKQLFCLDLYLHFITPLPEDHLTQPMFRGSSRSCLCEDPVCTPFVKKVF 242
Qy 145 SKQDVCKTLCTROPFSLAREACRSYSHVVLKEVRFENLQVLYPLSDPALNLRIVHLVRD 204
Db 243 EKHCKNRKCGPLNVTAAAEACRKEHMAKAVIRQLQLEFLQPLAEDPRLDLVQLVRD 302
Qy 205 PRAVLRSREAAGPILARDNGIVLG---TNGKWVE-----ADPHLRLIREVCRRSHVRIA 254
Db 303 PRAVLASRWVA-----PAGKYTKWKWLDDEGQDGLREEVQRLRGNCES-IRLS 351
Qy 255 EAATLKPPFLGRGRYLVRFEDLAREPLAEIRALYAFTGLTILTPOLEAHTHNTTH----G 310
Db 352 AELGLRQPAWLGRYMLVRYEDVARGPQKAREMYPFAGIPLTPQVEDWTQKNTQAADHG 411

```

Qy 311 SGIGKPIEAFHTSSRNARNVQAWRHLPFTKILRVQEVCAQALQLGYPVVSADQORD 370
 Db 412 SGI-----YSTQKSSQEFKWRFSMPFKLAQVQAPCGPAMRLFGYKLARDAAALTN 464
 Qy 371 LTLDLVLPGR 380
 Db 465 RSVSLBERG 474

RESULT 14
 US-08-655-878-2
 ; Sequence 2, Application US/08655878
 ; Patent No. 5827713
 ; GENERAL INFORMATION:
 ; APPLICANT: FUKUTA, MASAKAZU
 ; APPLICANT: HABUCHI, OSAMI
 ; TITLE OF INVENTION: DNA CODING FOR SULFOTRANSFERASE
 ; NUMBER OF SEQUENCES: 9
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE:
 ; STREET:
 ; CITY:
 ; STATE:
 ; COUNTRY:
 ; ZIP:
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/655,878
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME:
 ; REGISTRATION NUMBER:
 ; REFERENCE/DOCKET NUMBER:
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE:
 ; TELEFAX:
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 458
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 US-08-655-878-2

Query Match 26.4%; Score 540.5; DB 2; Length 458;
 Best Local Similarity 34.3%; Pred. No. 1.3e-50;
 Matches 124; Conservative 66; Mismatches 141; Indels 31; Gaps 8;
 Qy 36 AGGEDRVHVLVLSWRSGSSFLGQLFSQHPDVFLMEPAWIV--WTLTQCSAATLHMAV 93
 Db 107 AAPEFRHVLMTATRTGSSFVGFNQGNIFLFEPLWHIERTVTFEPGGANAVGSAL 166
 Qy 94 --RDLMSIFLCDMDVFDAYM---PQSRNLSAFFENWATSRALCSPPACSPAPRGITISKQD 148
 Db 167 VYRDVQLQLLCLDLYLESFISPAPEEHLTAALFRGSSHSLSCEPVTCTPSLKKVFEKYH 226
 Qy 149 VKTLCRTPQPSLAREACRSYSHVVLKEVRFNQLVPLPSDPALNLRIVHLVRDPRV 208
 Db 227 CKNRRCPLNITLAAEACRRKQHWALKTVIRQLEFLQPLAEDPRLDRIIQLVRDPRV 286
 Qy 209 LRSEAGPILARDNGIVLTNGKWE-----ADPHRLIREVCRSHVRIABATL 259
 Db 287 LVSRMVA-----FSGKYESKKWAABGEAPLQOEDEVQRLGNCEIS-IRLSAELGL 335
 Qy 260 KPPFFLGRVRLVRFEDLAREPLAEIRALYAFTGLTLPQLEAWIHNTHGSGIGKPIEA 319

Db 336 RQPRWLRGRYMLVRYEDVARAPLRKALEMYRFAGIHFTPOVEEWIRANTQAP---QDSNG 392
 Qy 320 FHTSSRNARNVQAWRHLPFTKILRVQEVCAQALQLGYPVVSADQORDLTLDLVLPR 379
 Db 393 IYSTQKSSQEFKWRFSIPFKLAQVQDACEPAMRLFGYKLASSAQELTNRSLSL-LEE 451
 Qy 380 GP 381
 Db 452 GP 453

RESULT 15
 US-09-015-188-2
 ; Sequence 2, Application US/09015188C
 ; Patent No. 6399358
 ; GENERAL INFORMATION:
 ; APPLICANT: Williams, Kevin J
 ; APPLICANT: Tabas, Ira
 ; TITLE OF INVENTION: A Human Gene Encoding Human Chondroitin
 ; TITLE OF INVENTION: 6-Sulfotransferase
 ; FILE REFERENCE: JEFF-0231
 ; CURRENT APPLICATION NUMBER: US/09/015,188C
 ; CURRENT FILING DATE: 1998-01-29
 ; NUMBER OF SEQ ID NOS: 17
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 2
 ; LENGTH: 411
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-015-188-2

Query Match 25.0%; Score 513.5; DB 3; Length 411;
 Best Local Similarity 33.5%; Pred. No. 1.1e-47;
 Matches 139; Conservative 63; Mismatches 170; Indels 43; Gaps 12;
 Qy 7 SSKTVTVLLLA---QTTCLLLFIISR---PGSSPAGGEDRV-----H 43
 Db 4 SWKAVLLALASIAIQVTAIRTTAKSFHTCPG-LAEAGLAERLCEBSPTPAYNLSRKTH 62
 Qy 44 VLVLSSWRSGSSFLGQLFSQHPDVFLMEPAWIVWTLT-----SQGSA-----ATLHMAVRD 95
 Db 63 ILILATTRSGSSFVGLFNQHLDFVLFEPFLYHVQNTLIPRTQKGSPADRRVMLGASRD 122
 Qy 96 LMBSIFLCDMDVFDAYM---PQSRNLSAFFENWATSRALCSPPACS-APRGITISKQDVCK 151
 Db 123 LRLSLYDCDLYFLENIYIKPPVNHDTTDRIFRGASRVLCSPVCDPPGADLVLEEGCV 182
 Qy 152 TLTCTROPFSLAREACRSYSHVVLKEVRFNQLVPLPSDPALNLRIVHLVRDPRVLR 211
 Db 183 RKCGLLNLTVAAEACRERSHVAIKTVRVPEVNDLRALVEDPRLNLKVQLVRDPRGILAS 242
 Qy 212 REAAGPILARDNGIVLTNGKWEADPHRLIREVCRSHVRIABATLKPPFPILGRYRL 271
 Db 243 RSETFRDTYRLWLWYTGKRPYNLD--VTQLTTVCEDFSNSVSTGLMR-PPWLKGYML 299
 Qy 272 VRDEDLAREPLAEIRALYAFTGLTLPQLEAWIHNTHGS-GIGKPIEAPHTSSRNARNV 330
 Db 300 VRDEDLARNPKKTEEYIGFLGIPLDSDHVARWVQNTNRGDPDTGK---HKYGTVRNSAT 356
 Qy 331 SQAWRHLPFTKILRVQEVCAQALQLGYPVVSADQORDLTLDLVRDPRGDFHS 385
 Db 357 AEKWRFLSYDIVAFAQNAQQVLAQLGYKIAASEBELKNPSVSLVEERDPRPFS 411

Search completed: June 23, 2005, 08:52:11
 Job time : 15.6998 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2005, 08:34:49 ; Search time 14.8883 Seconds
(without alignments)
1980.512 Million cell updates/sec

Title: US-10-697-828-7

Perfect score: 2057

Sequence: 1 MRLPRFSSTVMSLLMVQTG.....LPRGMSFKWASSTKQPES 395

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents AA.*
- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
 - 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
 - 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
 - 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
 - 5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
 - 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|----------------------|
| 1 | 1708 | 83.0 | 395 | 4 | US-09-949-016-7011 |
| 2 | 1536.5 | 74.7 | 390 | 4 | US-09-949-016-6813 |
| 3 | 1536.5 | 74.7 | 431 | 4 | US-09-949-016-8893 |
| 4 | 1017.5 | 49.5 | 386 | 3 | US-09-045-284A-2 |
| 5 | 1017.5 | 49.5 | 386 | 3 | US-09-130-911-1 |
| 6 | 986.5 | 48.0 | 386 | 4 | US-09-786-240-11 |
| 7 | 647.5 | 31.5 | 484 | 3 | US-09-263-023-4 |
| 8 | 647.5 | 31.5 | 484 | 4 | US-09-471-867-4 |
| 9 | 647.5 | 31.5 | 531 | 4 | US-09-949-016-6471 |
| 10 | 640 | 31.1 | 483 | 3 | US-09-263-023-2 |
| 11 | 640 | 31.1 | 483 | 4 | US-09-471-867-2 |
| 12 | 636.5 | 30.9 | 608 | 4 | US-09-949-016-9449 |
| 13 | 561.5 | 27.3 | 479 | 2 | US-08-899-514-2 |
| 14 | 549 | 26.7 | 411 | 3 | US-09-015-188-2 |
| 15 | 495 | 24.1 | 458 | 2 | US-08-655-878-2 |
| 16 | 96 | 4.7 | 468 | 4 | US-09-489-039A-14323 |
| 17 | 91 | 4.4 | 487 | 4 | US-09-252-931A-23907 |
| 18 | 90 | 4.4 | 933 | 1 | US-08-370-193A-8 |
| 19 | 90 | 4.4 | 933 | 3 | US-09-271-438A-9 |
| 20 | 90 | 4.4 | 933 | 4 | US-10-078-107-5 |
| 21 | 90 | 4.4 | 933 | 4 | US-10-077-751-5 |
| 22 | 89.5 | 4.4 | 294 | 4 | US-09-489-039A-12157 |
| 23 | 89.5 | 4.4 | 471 | 4 | US-09-252-931A-18083 |
| 24 | 89 | 4.3 | 421 | 4 | US-09-902-540-10244 |
| 25 | 88.5 | 4.3 | 268 | 4 | US-09-818-780-17 |
| 26 | 88.5 | 4.3 | 268 | 4 | US-09-818-780-94 |
| 27 | 88.5 | 4.3 | 561 | 4 | US-09-252-931A-24244 |

| | | | | | | |
|----|------|-----|------|---|----------------------|-------------------|
| 28 | 88.5 | 4.3 | 1395 | 2 | US-08-687-399-7 | Sequence 7, Appli |
| 29 | 88 | 4.3 | 675 | 4 | US-09-902-540-13407 | Sequence 13407, A |
| 30 | 87.5 | 4.3 | 575 | 4 | US-09-252-931A-32986 | Sequence 32986, A |
| 31 | 87.5 | 4.3 | 1098 | 3 | US-08-946-994-17 | Sequence 17, Appl |
| 32 | 87 | 4.2 | 984 | 2 | US-08-673-789-9 | Sequence 9, Appli |
| 33 | 85 | 4.1 | 938 | 4 | US-09-489-039A-13504 | Sequence 13504, A |
| 34 | 85 | 4.1 | 1299 | 5 | PCT-US95-08354A-2 | Sequence 2, Appli |
| 35 | 84.5 | 4.1 | 1338 | 4 | US-09-631-603-2 | Sequence 2, Appli |
| 36 | 84 | 4.1 | 496 | 3 | US-09-292-768-64 | Sequence 64, Appl |
| 37 | 84 | 4.1 | 984 | 2 | US-08-449-645A-19 | Sequence 19, Appl |
| 38 | 84 | 4.1 | 984 | 2 | US-08-702-367A-19 | Sequence 19, Appl |
| 39 | 84 | 4.1 | 984 | 5 | PCT-US95-04681-19 | Sequence 2, Appli |
| 40 | 84 | 4.1 | 1248 | 4 | US-10-042-810-2 | Sequence 2, Appli |
| 41 | 84 | 4.1 | 1278 | 4 | US-10-042-810-4 | Sequence 4, Appli |
| 42 | 83.5 | 4.1 | 460 | 4 | US-09-949-016-7994 | Sequence 7994, Ap |
| 43 | 83.5 | 4.1 | 671 | 4 | US-09-248-796A-17595 | Sequence 17595, A |
| 44 | 83.5 | 4.1 | 2867 | 4 | US-09-902-540-12593 | Sequence 12593, A |
| 45 | 83 | 4.0 | 474 | 4 | US-09-248-796A-17534 | Sequence 17534, A |

ALIGNMENTS

RESULT 1

US-09-949-016-7011

; Sequence 7011, Application US/09949016

; Patent No. 6812339

; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

; FILE REFERENCE: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 2000-04-14

; PRIOR APPLICATION NUMBER: 60/241,755

; PRIOR FILING DATE: 2000-10-20

; PRIOR APPLICATION NUMBER: 60/237,768

; PRIOR FILING DATE: 2000-10-03

; PRIOR APPLICATION NUMBER: 60/231,498

; PRIOR FILING DATE: 2000-09-08

; NUMBER OF SEQ ID NOS: 207012

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 7011

; LENGTH: 395

; TYPE: PRT

; ORGANISM: Human

US-09-949-016-7011

Query Match 83.0%; Score 1708; DB 4; Length 395;

Best Local Similarity 83.0%; Pred. No. 1.4e-181;

Matches 328; Conservative 24; Mismatches 43; Indels 0; Gaps 0;

| | | | |
|----|-----|--|-----|
| Qy | 1 | MRLPRFSSTVMSLLMVQTGILVLSRQVPSPPAGLGERVHVLSWSRSGSFGVGLF | 60 |
| Db | 1 | MMLPRVSTAVTALLAQTELLFLVSRPGSGAGEARVHVLSWSRSGSFGVGLF | 60 |
| Qy | 61 | SOQPDVFLMEPAHWVMDTLSQGSAPALHMAVRLIRSVFLCDMDVDAYLPMRNI | 120 |
| Db | 61 | NQHPDVFYLMPEPAHWVMDTLSQGSAAHLMAVRLIRSVFLCDMDVDAYLPMRNI | 120 |
| Qy | 121 | FOVAVSRALCSPVCEAFARGNISSEBCKPLCATRPFGLAQACSSYSHVVLKEV | 180 |
| Db | 121 | FOVAVSRALCSPVCEAFARGNISSEBCKPLCATRPFGLAQACSSYSHVVLKEV | 180 |
| Qy | 181 | LOVLYPLSDPALNRIHVLRDPRVLRREQTAKALARDNGIVLGTNGTWEADPLR | 240 |
| Db | 181 | LOVLYPLSDPALNRIHVLRDPRVLRREQTAKALARDNGIVLGTNGTWEADPLR | 240 |
| Qy | 241 | VVNEVCBSHVRIAEALHKPPPLQDRYLVRVEDLARDPLTVIRELYAFTGLT | 300 |
| Db | 241 | VVNEVCBSHVRIAEALHKPPPLQDRYLVRVEDLARDPLTVIRELYAFTGLT | 300 |
| Qy | 301 | TIWNITHGSGPGARREAFKTTSRDALSVSQAWRHLPFAKIRRVQELCGGALQ | 360 |

```

; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8893
; LENGTH: 390
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6813

Query Match 74.7%; Score 1536.5; DB 4; Length 390;
Best Local Similarity 76.0%; Pred. No. 1.9e-162;
Matches 295; Conservative 30; Mismatches 62; Indels 1; Gaps 1;

Db 301 AWIHNTHGSGPGARREAFKTSRRNALNSQAWHEALPFAKIRRVQELCAGALQLLGYR 360
Qy 361 VHSLEQRDLSDLLPRGMDSEFKWASSTKQPS 395
Db 361 VYSDEQRNALDLVLRGLNGFTWASSTASHPRN 395

RESULT 2
US-09-949-016-6813
; Sequence 6813, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8813
; LENGTH: 390
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6813

Query Match 74.7%; Score 1536.5; DB 4; Length 390;
Best Local Similarity 76.0%; Pred. No. 1.9e-162;
Matches 295; Conservative 30; Mismatches 62; Indels 1; Gaps 1;

Qy 1 MRLPRFSTVWLSLMLVOTG-ILVFLVSRQVPSPAGLGERVHVLSLSRSGSSFGQL 59
Db 1 MRLPRFSTVWLSLMLVOTG-ILVFLVSRQVPSPAGLGERVHVLSLSRSGSSFGQL 59
Qy 60 FSQHPDVFLMEPAWHVWDTLSQGSAPALHMAVRDLIRSVFLCDMDVDFDAYLPWRNID 119
Db 61 FSQHPDVFLMEPAWHVWDTLSQGSAPALHMAVRDLIRSVFLCDMDVDFDAYLPWRNID 119
Qy 120 LFQWAVSRALCSPPVCEAFARGNISSEVCKPLCATRPFGLAQEACSSYSHVVVKVRFF 179
Db 121 FFWWATSRALCSPPVCEAFARGNISSEVCKPLCATRPFGLAQEACSSYSHVVVKVRFF 179
Qy 180 NLQVLYPLSDPALNLRIVHLVRDPAVLRSREQTAKALARDNGIVLGTNGTWVEADPRL 239
Db 181 NLQVLYPLSDPALNLRIVHLVRDPAVLRSREQTAKALARDNGIVLGTNGTWVEADPRL 239
Qy 240 RVNVECRSHVRIAEALHKKPPPLQDRYRLRYVEDLARDPLTVIRELYAFTGLTLPOL 299
Db 241 RLIREVCRSHVRIAEALHKKPPPLQDRYRLRYVEDLARDPLTVIRELYAFTGLTLPOL 299
Qy 300 QTWIHNTHGSGPGARREAFKTSRRDALSVQAWRHHTLPFAKIRRVQELCGGALQLLGYR 359
Db 301 EAWIHNTHGSGIGKPIEAFTSSRNARNVQAWRHHTLPFAKIRRVQELCGGALQLLGYR 360
Qy 360 SVHSELRDLSDLLPRGMDSEFKWAS 387
Db 361 PVYSADQQRDLTLDLVLPRGPDHFSWAS 388

RESULT 3
US-09-949-016-8893
; Sequence 8893, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

```

```

; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8893
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-8893

Query Match 74.7%; Score 1536.5; DB 4; Length 431;
Best Local Similarity 76.0%; Pred. No. 2.2e-162;
Matches 295; Conservative 30; Mismatches 62; Indels 1; Gaps 1;

Qy 1 MRLPRFSTVWLSLMLVOTG-ILVFLVSRQVPSPAGLGERVHVLSLSRSGSSFGQL 59
Db 42 MMLPRFSSKTVTVLLLAQTCLLLFIISRPSPAGGEDRVHVLVLSRSGSSFLGQL 101
Qy 60 FSQHPDVFLMEPAWHVWDTLSQGSAPALHMAVRDLIRSVFLCDMDVDFDAYLPWRNID 119
Db 102 FSQHPDVFLMEPAWHVWDTLSQGSAPALHMAVRDLIRSVFLCDMDVDFDAYLPWRNID 119
Qy 120 LFQWAVSRALCSPPVCEAFARGNISSEVCKPLCATRPFGLAQEACSSYSHVVVKVRFF 179
Db 162 FFWWATSRALCSPPVCEAFARGNISSEVCKPLCATRPFGLAQEACSSYSHVVVKVRFF 179
Qy 180 NLQVLYPLSDPALNLRIVHLVRDPAVLRSREQTAKALARDNGIVLGTNGTWVEADPRL 239
Db 222 NLQVLYPLSDPALNLRIVHLVRDPAVLRSREQTAKALARDNGIVLGTNGTWVEADPRL 239
Qy 240 RVNVECRSHVRIAEALHKKPPPLQDRYRLRYVEDLARDPLTVIRELYAFTGLTLPOL 299
Db 282 RLIREVCRSHVRIAEALHKKPPPLQDRYRLRYVEDLARDPLTVIRELYAFTGLTLPOL 299
Qy 300 QTWIHNTHGSGPGARREAFKTSRRDALSVQAWRHHTLPFAKIRRVQELCGGALQLLGYR 359
Db 342 EAWIHNTHGSGIGKPIEAFTSSRNARNVQAWRHHTLPFAKIRRVQELCGGALQLLGYR 401
Qy 360 SVHSELRDLSDLLPRGMDSEFKWAS 387
Db 402 PVYSADQQRDLTLDLVLPRGPDHFSWAS 429

RESULT 4
US-09-045-284A-2
; Sequence 2, Application US/09045284A
; Patent No. 6265192
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107U51
; CURRENT APPLICATION NUMBER: US/09/045,284A
; CURRENT FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-045-284A-2

Query Match 49.5%; Score 1017.5; DB 3; Length 386;
Best Local Similarity 56.4%; Pred. No. 1.5e-104;

```

Matches 215; Conservative 43; Mismatches 112; Indels 11; Gaps 7;

Qy 1 MRLPRFSSTVMSLMLVQTGILVF---LVSRQVPS-SPAGLGERVHVHLVLSWSRSGSFV 56
 Db 1 MLLPK--KMKLLFLVVSQMAILALFFHMYSHNISLSMKQAQPERMHVHLVLSWSRSGSFV 58
 Qy 57 GOLFSQHPDVFYLMPEPAWHVMDTSLQSGAPALHMAVRDLIRSVFLCDMDVFDAYL-PWRR 115
 Db 59 GOLFGQHPDVFYLMPEPAWHVMDTSLQSGAPALHMAVRDLIRSVFLCDMDVFDAYMEPGPR 118
 Qy 116 NISDLFQWAVSRALCSPVCEAFARGNISSEVCKPLCATRPFGLAQEACSSYSHVVLKE 175
 Db 119 RQSSLFQWNSRALCSAPACDIIPQDEIIPRAHCRLLCSQOPFEVVEKACRSYSHVVLKE 178
 Qy 176 VRFFNLQVLYPLSDPALNLRIVHLVRDPRAVLSRREQTAKALARDNGIVLGTNGTWV-E 234
 Db 179 VRFFNLQSLYPLKDPNSLNHLVHLVRDPRAVLSRREQTAKALARDNGIVLGTNGTWV-E 238
 Qy 235 ADPRLRVNVECRSHVRIAEALHKKPPFPLODRYLRVRYEDLARDPLTVIRELYAFTGLG 294
 Db 239 EDQPYVMQVICSQLEIYK-TIQSLPKALQERYLLVRYEDLARAPVAQTSRMVEFVGL 297
 Qy 295 LTPOLQVTHNITHGSGPGARREAFKTTSDALSVSQAWRHTLPFAKIRRVQELCGGALQ 354
 Db 298 FLPHLQVTHNITHGSGPGARREAFKTTSDALSVSQAWRHTLPFAKIRRVQELCGGALQ 354
 Qy 355 LLGYRSVHSELEQDLSLDLL 375
 Db 356 LLGYRHVRSEQRNLLDILL 376

RESULT 5

US-09-190-911-1
 ; Sequence 1, Application US/09190911
 ; Patent No. 6365365
 ; GENERAL INFORMATION:
 ; APPLICANT: Bistrup, Annette
 ; APPLICANT: Rosen, Steven D.
 ; APPLICANT: Tangemann, Kirsten
 ; APPLICANT: Hemmerich, Stefan
 ; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
 ; FILE REFERENCE: 6510-107CIP
 ; CURRENT APPLICATION NUMBER: US/09/190,911
 ; EARLIER FILING DATE: 1998-11-12
 ; EARLIER APPLICATION NUMBER: 09/045,284
 ; NUMBER OF SEQ ID NOS: 8
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 1
 ; LENGTH: 386
 ; TYPE: PRT
 ; ORGANISM: H. sapiens
 US-09-190-911-1

Query Match 49.5%; Score 1017.5; DB 3; Length 386;
 Best Local Similarity 56.4%; Pred. No. 1.5e-104;
 Matches 215; Conservative 43; Mismatches 112; Indels 11; Gaps 7;

Qy 1 MRLPRFSSTVMSLMLVQTGILVF---LVSRQVPS-SPAGLGERVHVHLVLSWSRSGSFV 56
 Db 1 MLLPK--KMKLLFLVVSQMAILALFFHMYSHNISLSMKQAQPERMHVHLVLSWSRSGSFV 58
 Qy 57 GOLFSQHPDVFYLMPEPAWHVMDTSLQSGAPALHMAVRDLIRSVFLCDMDVFDAYL-PWRR 115
 Db 59 GOLFGQHPDVFYLMPEPAWHVMDTSLQSGAPALHMAVRDLIRSVFLCDMDVFDAYMEPGPR 118
 Qy 116 NISDLFQWAVSRALCSPVCEAFARGNISSEVCKPLCATRPFGLAQEACSSYSHVVLKE 175
 Db 119 RQSSLFQWNSRALCSAPACDIIPQDEIIPRAHCRLLCSQOPFEVVEKACRSYSHVVLKE 178
 Qy 176 VRFFNLQVLYPLSDPALNLRIVHLVRDPRAVLSRREQTAKALARDNGIVLGTNGTWV-E 234
 Db 179 VRFFNLQSLYPLKDPNSLNHLVHLVRDPRAVLSRREQTAKALARDNGIVLGTNGTWV-E 238

Qy 235 ADPRLRVNVECRSHVRIAEALHKKPPFPLODRYLRVRYEDLARDPLTVIRELYAFTGLG 294
 Db 239 EDQPYVMQVICSQLEIYK-TIQSLPKALQERYLLVRYEDLARAPVAQTSRMVEFVGL 297
 Qy 295 LTPOLQVTHNITHGSGPGARREAFKTTSDALSVSQAWRHTLPFAKIRRVQELCGGALQ 354
 Db 298 FLPHLQVTHNITHGSGPGARREAFKTTSDALSVSQAWRHTLPFAKIRRVQELCGGALQ 354
 Qy 355 LLGYRSVHSELEQDLSLDLL 375
 Db 356 LLGYRHVRSEQRNLLDILL 376

RESULT 6

US-09-786-240-11
 ; Sequence 11, Application US/09786240
 ; Patent No. 6558935
 ; GENERAL INFORMATION:
 ; APPLICANT: INCYTE PHARMACEUTICALS, INC.
 ; APPLICANT: TANG, Y. Tom
 ; APPLICANT: CORLEY, Neil C.
 ; APPLICANT: GUEGLER, Karl J.
 ; APPLICANT: BAUGHN, Mariah R.
 ; APPLICANT: LAL, Preeti
 ; APPLICANT: YUE, Henry
 ; APPLICANT: HILLMAN, Jennifer L.
 ; APPLICANT: AZIMZAI, Valda
 ; TITLE OF INVENTION: HUMAN TRANSFERASE PROTEINS
 ; FILE REFERENCE: PF-0592 PCT
 ; CURRENT APPLICATION NUMBER: US/09/786,240
 ; CURRENT FILING DATE: 2002-03-12
 ; PRIOR APPLICATION NUMBER: 09/150,657; unassigned; 09/186,779; unassigned; 60/133,642
 ; PRIOR FILING DATE: 1998-09-10; 1998-09-10; 1998-11-04; 1998-11-04; 1999-05-11
 ; NUMBER OF SEQ ID NOS: 33
 ; SOFTWARE: PERL Program
 ; SEQ ID NO 11
 ; LENGTH: 386
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; OTHER INFORMATION: Incyte ID No. 6558935 2617407CD1
 US-09-786-240-11

Query Match 48.0%; Score 986.5; DB 4; Length 386;
 Best Local Similarity 56.3%; Pred. No. 4.2e-101;
 Matches 215; Conservative 40; Mismatches 114; Indels 13; Gaps 9;

Qy 1 MRLPRFSSTVMSLMLVQTGILVF---LVSRQVPS-SPAGLGERVHVHLVLSWSRSGSFV 56
 Db 1 MLLPK--KMKLLFLVVSQMAILALFFHMYSHNISLSMKQAQPERMHVHLVLSWSRSGSFV 58
 Qy 57 GOLFSQHPDVFYLMPEPAWHVMDTSLQSGAPALHMAVRDLIRSVFLCDMDVFDAYL-PWRR 115
 Db 59 GOLFGQHPDVFYLMPEPAWHVMDTSLQSGAPALHMAVRDLIRSVFLCDMDVFDAYMEPGPR 118
 Qy 116 NISDLFQWAVSRALCSPVCEAFARGNISSEVCKPLCATR-PFGLAQEACSSYSHVVLK 174
 Db 119 RQSSLFQWNSRALCSAPACDIIPQDE-SSFGLTAGCAVNSPLKLEKACRSYSHVVLK 177
 Qy 175 VRFFNLQVLYPLSDPALNLRIVHLVRDPRAVLSRREQTAKALARDNGIVLGTNGTWV- 233
 Db 178 VRFFNLQSLYPLKDPNSLNHLVHLVRDPRAVLSRREQTAKALARDNGIVLGTNGTWV- 237
 Qy 234 EADPRLRVNVECRSHVRIAEALHKKPPFPLODRYLRVRYEDLARDPLTVIRELYAFTGL 293
 Db 238 KEDQPYVMQVICSQLEIYK-TIQSLPKALQERYLLVRYEDLARAPVAQTSRMVEFVGL 296
 Qy 294 GLTQOLQVTHNITHGSGPGARREAFKTTSDALSVSQAWRHTLPFAKIRRVQELCGGAL 353
 Db 297 EFLPHLQVTHNITHGSGPGARREAFKTTSDALSVSQAWRHTLPFAKIRRVQELCGGAL 354

Qy 354 QLLGYSVHSELEQDLSLDLL 375
Db 355 NLLGYSVHSELEQDLSLDLL 376

RESULT 7

US-09-263-023-4
; Sequence 4, Application US/09263023
; Patent No. 6037159
; GENERAL INFORMATION:
; APPLICANT: Uchimura, Kenji
; APPLICANT: Muramatsu, Hideki
; APPLICANT: Kadomatsu, Kenji
; APPLICANT: Kannagi, Reiji
; APPLICANT: Habuchi, Osami
; APPLICANT: Muramatsu, Takashi
; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-SULFOTRANSFERASE AND
; FILE REFERENCE: TOYAM41.001AUS
; CURRENT APPLICATION NUMBER: US/09/263,023
; CURRENT FILING DATE: 1999-03-05
; EARLIER APPLICATION NUMBER: JP 10-54007
; EARLIER FILING DATE: 1998-03-05
; EARLIER APPLICATION NUMBER: JP 10-177844
; EARLIER FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-263-023-4

Query Match 31.5%; Score 647.5; DB 3; Length 484;
Best Local Similarity 36.8%; Pred. No. 4e-63;
Matches 140; Conservative 76; Mismatches 117; Indels 47; Gaps 9;

Qy 33 SPAGLGERVH-VLVLSWRSGSFVQGLFSQHPDVFLMEPAWHVWDTLSQGSAPALHMA 91
Db 110 APEGVGDKRHMYVFTTWRSGSFGELEFNQNPVEFFLYEPVHVMQKLYPGDAVSLQGA 169
Qy 92 VRDLRSVFLCDMDVFDAYLP---WRRNISDL--FQWAVSRALCSPPVCEAFARG--NIS 144
Db 170 ARDMSALYRCDSLVSFQLYSPAGSGGRNLTGLFGAATNKVVCSSPLCPAYRKEVUGLV 229
Qy 145 SEEVCKPLCATRPPGLAQEACSSYSHVVLKEVRFNNQVLYPLSDPALNLRIVHLVRDP 204
Db 230 DDRVCKK-CPQRLARFECECRKYRTLVIKGVRFVDVAVLAPLLRDPALDPLKVIHLVRDP 288
Qy 205 RAVLSRREQTAKALARDNGIVLGTNGTWVEADPR----- 238
Db 289 RAVASSRIRSRHGLIRSLQVVRSR-----DPAHRMPFLEAAGHKLGAKEGVGGPAD 342
Qy 239 ---LRVNVNCSHVRIAEALHKPPFLQDRVRLVRYEDLARDPLTVIRELYAFTGLGL 295
Db 343 YHALGAMEVICNSMAKTLOTAL-QPPDWLQGHYLVVRYEDLVGDPVKTLRRVYDFVGLLV 401
Qy 296 TPQLQTWIHNITHGSGPGARREAFKTTSRDALSVSQAWRHTLPFAKIRRVQELCGGALQL 355
Db 402 SPEMEQFALNMTSGSGSSK--PFVVSARNATQAAANWRTALTFFQIKQVEEFCYQPMV 459
Qy 356 LGYRSVHSELEQDLSLDLL 375
Db 460 LGYRVNSPEEVKDLSTKLL 479

RESULT 8

US-09-471-867-4
; Sequence 4, Application US/09471867
; Patent No. 6455289
; GENERAL INFORMATION:
; APPLICANT: Uchimura, Kenji
; APPLICANT: Muramatsu, Hideki

; APPLICANT: Kadomatsu, Kenji
; APPLICANT: Kannagi, Reiji
; APPLICANT: Habuchi, Osami
; APPLICANT: Muramatsu, Takashi
; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-SULFOTRANSFERASE AND
; FILE REFERENCE: TOYAM41.001AUS
; CURRENT APPLICATION NUMBER: US/09/471,867
; CURRENT FILING DATE: 1999-12-23
; EARLIER APPLICATION NUMBER: US 09/263,023
; EARLIER FILING DATE: 1999-03-05
; EARLIER APPLICATION NUMBER: JP 10-54007
; EARLIER FILING DATE: 1998-03-05
; EARLIER APPLICATION NUMBER: JP 10-177844
; EARLIER FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-471-867-4

Query Match 31.5%; Score 647.5; DB 4; Length 484;
Best Local Similarity 36.8%; Pred. No. 4e-63;
Matches 140; Conservative 76; Mismatches 117; Indels 47; Gaps 9;

Qy 33 SPAGLGERVH-VLVLSWRSGSFVQGLFSQHPDVFLMEPAWHVWDTLSQGSAPALHMA 91
Db 110 APEGVGDKRHMYVFTTWRSGSFGELEFNQNPVEFFLYEPVHVMQKLYPGDAVSLQGA 169
Qy 92 VRDLRSVFLCDMDVFDAYLP---WRRNISDL--FQWAVSRALCSPPVCEAFARG--NIS 144
Db 170 ARDMSALYRCDSLVSFQLYSPAGSGGRNLTGLFGAATNKVVCSSPLCPAYRKEVUGLV 229
Qy 145 SEEVCKPLCATRPPGLAQEACSSYSHVVLKEVRFNNQVLYPLSDPALNLRIVHLVRDP 204
Db 230 DDRVCKK-CPQRLARFECECRKYRTLVIKGVRFVDVAVLAPLLRDPALDPLKVIHLVRDP 288
Qy 205 RAVLSRREQTAKALARDNGIVLGTNGTWVEADPR----- 238
Db 289 RAVASSRIRSRHGLIRSLQVVRSR-----DPAHRMPFLEAAGHKLGAKEGVGGPAD 342
Qy 239 ---LRVNVNCSHVRIAEALHKPPFLQDRVRLVRYEDLARDPLTVIRELYAFTGLGL 295
Db 343 YHALGAMEVICNSMAKTLOTAL-QPPDWLQGHYLVVRYEDLVGDPVKTLRRVYDFVGLLV 401
Qy 296 TPQLQTWIHNITHGSGPGARREAFKTTSRDALSVSQAWRHTLPFAKIRRVQELCGGALQL 355
Db 402 SPEMEQFALNMTSGSGSSK--PFVVSARNATQAAANWRTALTFFQIKQVEEFCYQPMV 459
Qy 356 LGYRSVHSELEQDLSLDLL 375
Db 460 LGYRVNSPEEVKDLSTKLL 479

RESULT 9

US-09-949-016-6471
; Sequence 6471, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08

Db 209 KVCSSPLCPAYRKEVVGLVDVDRCKK-CPPQRLARFEEBCRYRTVVIKGVRFVDFVAVL 267
QY 185 YELSDPALNLRIVHLVRDRAVLRSQEQAKALARDNGIVLGTNGTWVEADPR----- 238
Db 268 APLKOPALDOLKVIHLVRDPRVASSRIRSRKHGLIRESLQVRSR-----DPAHRMPF 321
QY 239 -----LRVNEVCRSHVRIAEALHKPPPPFLQDRYLRVRYED 275
Db 322 LEAAGHKLGAKEGMPGADYHALGAMEVICNSMAKTLQAL-QPPDWLQGHYLVRYED 360
QY 276 LARDPLTVIRELYAFTGLTGPQLQVWIHNITGSGPGARREAFKTTSDALSVSQAWRH 335
Db 381 LVGDPVKTLRRVYDFVGLLVSPMEQPALNMTSGSSSK--PFVWSARNATQANAWRT 438
QY 336 TLPKATRRVQELCGGALQLLGLVRSVHSELEQRDLSLDLL 375
Db 439 ALTFQKIQVEEFCYQPMALVGLVRSVNSPEEVKDLSTLL 478
RESULT 12
US-09-949-016-9449
; Sequence 9449, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CU001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9449
; LENGTH: 608
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-9449
Query Match 30.9%; Score 636.5; DB 4; Length 608;
Best Local Similarity 37.1%; Pred. No. 9.8e-62;
Matches 140; Conservative 73; Mismatches 117; Indels 47; Gaps 9;
QY 36 GLGE-RVHVLVLSWSRSSGSSFGVQLFSQHPDVFYLMPEPAWHVMDTISQGSAPALHMAVRD 94
Db 237 GVGDKRLVVFTWRSSSFFGELFNQNPVEFVYEPVHVWQKLYPGDAVSIQGNARD 236
QY 95 LIRSVFLCDMDVFDAYLP---WRRNISDL--FQWAVSRALCSPVCEAFARG--NTISSE 147
Db 297 MLGALYRCDSLVSFOLYSPAGSGGRNLTGLFGAATNKVVCSSPLCPAYRKEVVGLVDDR 356
QY 148 VCKPLCATRPFGLAQEACSSYSHVVLKEVFFNLQVLYPLSDPALNLRIVHLVRDPRV 207
Db 357 VCKK-CPQRLARFEEBCRYRTLVIKGVRFVDFVAVLAPLRLDPALDVKVIHLVRDPRV 415
QY 208 LRSEQAKALARDNGIVLGTNGTWVEADPR----- 238
Db 416 ASSRIRSHGLIRESLQVRSR-----DPAHRMPFLEAGHKLGAKEGVCGPADYHA 469
QY 239 LRVNVEVCRSHVRIAEALHKPPPPFLQDRYLRVRYEDLARDPLTVIRELYAFTGLGTTPQ 298
Db 470 LGAMEVICNSMAKTLQAL-QPPDWLQGHYLVRYEDLVGDPVKTLRRVYDFVGLLVSP 528
QY 299 LQVWIHNITGSGPGARREAFKTTSDALSVSQAWRHTLPFAKIRRVQELCGGALQILGY 358
Db 529 MEQPALNMTSGSSSK--PFVWSARNATQANAWRTALTFFQIKQVEEFCYQPMALVGY 586
QY 359 RSVHSELEQRDLSLDLL 375

Db 587 ERVNSPEEVKDLSTLL 603
RESULT 13
US-08-899-514-2
; Sequence 2, Application US/08899514
; Patent No. 5910581
; GENERAL INFORMATION:
; APPLICANT: HABUCHI, OSAMI
; APPLICANT: FUKUCHI, MASAKAZU
; TITLE OF INVENTION: POLYPEPTIDE OF GLYCOSAMINOGLYCAN
; TITLE OF INVENTION: SULFOTRANSFERASE ORIGINATING FROM HUMAN AND DNA CODING
; TITLE OF INVENTION: FOR THE POLYPEPTIDE
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: KNOBBE, MARTENS, OLSON & BEAR, LLP
; STREET: 620 NEWPORT CENTER DRIVE, SIXTEENTH FLOOR
; CITY: NEWPORT BEACH
; STATE: CALIFORNIA
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/899,514
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: DANIEL E ALTMAN
; REGISTRATION NUMBER: 34,115
; REFERENCE/DOCKET NUMBER: TOYAM21.001AUS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 714 760 0404
; TELEFAX: 714 760 9502
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 479
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-899-514-2
Query Match 27.3%; Score 561.5; DB 2; Length 479;
Best Local Similarity 38.6%; Pred. No. 1.6e-53;
Matches 142; Conservative 55; Mismatches 136; Indels 35; Gaps 11;
QY 31 PPSAGLGRVHRVHLVLSWSRSSGSSFGVQLFSQHPDVFYLMPEPAWHVMDTIS--QGSAPAL 88
Db 123 PPRPAVAGPRRHVLLMATTTRTSSFGVEFFNQGNIFYLFEPLWHITRTVSPFPGANAA 182
QY 89 HMAV--RDLIRSVFLCDMDVFDAY---LPWRRNISDLFQWAVSRALCSPVCEAFARGNI 143
Db 183 GSALVYRDVLKQLFLCDLYLVEHFTPLPEDLHTLQPMFRGSSRSRLCEDPVCCTPFVK-KV 241
QY 144 SSEEVCCK----PLCATRPFGLAQEACSSYSHVVLKEVFFNLQVLYPLSDPALNLRIV 198
Db 242 FEKYHCNRCRCPGLNVT-----LAAEACRKEHMAKAVIRQLFELOPLAEDPRDLRVI 297
QY 199 HLVRDPRVLRSR-----EQAKALARDNGIVLGTNGTWVEADPRLRVNEVCRSHVR 251
Db 298 QLVRDPRVLRSMWAFAGKYTKWKWLDDE---GQDGLREEEVQRLR---GNCESIRL 350
QY 252 IAEALHKPPPPFLQDRYLRVRYEDLARDPLTVIRELYAFTGLTLPOLQVWIHNITGSG 311
Db 351 SAEGLG-RQPALRGYMLRVYEDVARGPLQKAREMYPFAGIPLTTPQVEDWIQNTQAAH 409

Qy 312 PGARREAFKTTSDALSVQAWHTLPPAKIRRVQELCGGALQQLGYRSVHSELEQRDLS 371
Db 410 DGS---GIYTKNSSEQEFKWRFSMPFKLAQVVQAPCGPAMRFLGYKLARDAALNRS 466
Qy 372 LDLLLPRG 379
Db 467 VSLLEERG 474

RESULT 14

US-09-015-188-2
; Sequence 2, Application US/09015188C
; Patent No. 6399358
; GENERAL INFORMATION:
; APPLICANT: Williams, Kevin J
; APPLICANT: Tabas, Ira
; TITLE OF INVENTION: A Human Gene Encoding Human Chondroitin
; TITLE OF INVENTION: 6-Sulfotransferase
; FILE REFERENCE: JEFF-0231
; CURRENT APPLICATION NUMBER: US/09/015.188C
; CURRENT FILING DATE: 1998-01-29
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 411
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-015-188-2

Query Match 26.7%; Score 549; DB 3; Length 411;
Best Local Similarity 36.4%; Pred. No. 3e-52;
Matches 136; Conservative 58; Mismatches 160; Indels 20; Gaps 9;

Qy 25 LVSRQVSSPA---GLGERVHVLSSWRSGSVFGQLFSQHPDVFYLMPEAHVMDTL- 80
Db 42 LAERLCESPTFAYNLSRKTHILILATRSQSVFGQLFNQHLDFYLFEPYHVQNTLI 101
Qy 81 ---SQGSAPA---LHMAVRLIRSVFLCDMDVFDAYL---PWRNISDLFQWAVSRALC 130
Db 102 PRFTQGSPPADRRVYMLGASRLLSLYCDLYFLENYIKPPVNHVHTDRIFRGASVLC 161
Qy 131 SPVCEAFARNISSEE-VCKPLCATRPFGLAQACSSYSHVVLKEVRFNQLVYPLLS 189
Db 162 SRPVCDDPGPADLVLEEGDCVRKCGLLNLTVAAEACRSHRSHVIAKTVRVPVNDLRALVE 221
Qy 190 DPALNLRVHVRPRAVLRSEOTAKALARDNGIVLTNGTWVADPRLRVNVNVCESH 249
Db 222 DPRLNLKVQLVRDPRGILASRSETFRDYLRLWLYGTGRKPYNLD--VTQLTTVCEDF 279
Qy 250 VRIEAAALHKPPPLQDRYRLVRVEDLARDPLTVIRELYAFTGLTLPQLQTIHNIHG 309
Db 280 SNSYSTGLMR-PPWLKGYMLVRVEDLARNPKMTEIYIGFLIPLDSHVARNIQNTRG 338
Qy 310 SGPGARREAFKTTSDALSVQAWHTLPPAKIRRVQELCGGALQQLGYRSVHSELEQRD 369
Db 339 D-PTLGKHKYCTV-RNSAATAEKWRFLSYDIVAFAQACQVLAQLGYKIAASEELKN 396
Qy 370 LSLDLLPRGMSDF 383
Db 397 PSVSLVEERDPRFP 410

RESULT 15

US-08-655-878-2
; Sequence 2, Application US/08655878
; Patent No. 5827713
; GENERAL INFORMATION:
; APPLICANT: FUKUTA, MASAKAZU
; APPLICANT: HABUCHI, OSAMI
; TITLE OF INVENTION: DNA CODING FOR SULFOTRANSFERASE
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE:

STREET:
CITY:
STATE:
COUNTRY:
ZIP:
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/655,878
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME:
REGISTRATION NUMBER:
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE:
TELEFAX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 458
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-655-878-2

Query Match 24.1%; Score 495; DB 2; Length 458;
Best Local Similarity 36.1%; Pred. No. 3.8e-46;
Matches 129; Conservative 59; Mismatches 125; Indels 44; Gaps 13;

Qy 40 RVHVLSSWRSGSVFGQLFSQHPDVFYLMPEAHVMDTL---QGSAPALHMAV--RDL 95
Db 112 RRHVLMMATRTGSSFVGBFNFQGNIFYLFEPLMHIERTVTFPGGANAVGSALVYRDV 171
Qy 96 IRSVFLCDMDVFDAYL---PWRNISDLFQWAVSRALCSPVC-----EAFARNISSEE 147
Db 172 LQQLLLCDLYLESFISFSPAPEEHLTAALFRGSSHSLCEBPVCTPSLKKVFEKTHCKNRR 231
Qy 148 VCKPLCATRPFGLAQACSSYSHVVLKEVRFNQLVYPLLSDPALNLRIVHLVRDPRAV 207
Db 232 -CGPLNIT-----LAAEACRRKQHWALKTVIRQLFLOPLAEDPRLDLRIQLVRDPRAV 286
Qy 208 LRSREOTAKALARDNGIVLTNGTW-----VEADPRLRVNVNEV-----CRSHVRIABAL 257
Db 287 LVSRMVAFS-----GKYESWKWAAEAGEAPLQ-EDEVQRLRGNCESIRLSAELGL 335
Qy 258 HKPPPLQDRYRLVRVEDLARDPLTVIRELYAFTGLTLPQLQTIHNIHGSGPGARRE 317
Db 336 -RQPRWLRGMYLVRVEDVARAPLRKALEMYRFAGIHPTFQVEEMIRANTQAPQDS---N 391
Qy 318 AFTTSDALSVQAWHTLPPAKIRRVQELCGGALQQLGYRSVHSELEQRDLSL 372
Db 392 GIYSTQKNSSEQEFKWRFSIPFKLAQVVQAPCGPAMRFLGYKLARDAALNRSLSL 448

Search completed: June 23, 2005, 08:52:10
Job time : 16.8883 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2005, 08:32:43 ; Search time 64.4109 Seconds
(without alignments)
3578.727 Million cell updates/sec

Title: US-10-697-828-9

Perfect score: 3147

Sequence: 1 MNRINGAMVDVDAHYKMFV.....LIENICWTLMRLGLGPKFMD 596

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : A_Geneseq_16Dec04:*

- 1: Geneseq1980s:*
- 2: Geneseq1990s:*
- 3: Geneseq2000s:*
- 4: Geneseq2001s:*
- 5: Geneseq2002s:*
- 6: Geneseq2003as:*
- 7: Geneseq2003bs:*
- 8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|--------------------|
| 1 | 3147 | 100.0 | 596 | 4 | AAY72641 Human gly |
| 2 | 3147 | 100.0 | 1222 | 4 | AAY72642 Human gly |
| 3 | 3142 | 99.8 | 1212 | 6 | AAE33541 Human nov |
| 4 | 3142 | 99.8 | 1222 | 6 | ABU11849 Human sec |
| 5 | 2835 | 90.1 | 1207 | 4 | AAY72643 Mouse gly |
| 6 | 2835 | 90.1 | 1207 | 6 | AAE33542 Human nov |
| 7 | 1821 | 57.9 | 755 | 4 | AAE33543 Human pro |
| 8 | 1821 | 57.9 | 755 | 8 | ADP56673 Human pro |
| 9 | 920 | 29.2 | 480 | 4 | AAE33544 Human pro |
| 10 | 637 | 20.2 | 128 | 4 | ADG27796 Human nov |
| 11 | 615 | 19.5 | 314 | 4 | AAE33545 Human pro |
| 12 | 568 | 18.0 | 125 | 8 | ADJ12217 Human sec |
| 13 | 568 | 18.0 | 126 | 2 | AAE33546 Human sec |
| 14 | 147 | 4.7 | 410 | 7 | ADN95522 Human BEC |
| 15 | 147 | 4.7 | 411 | 2 | AAE33547 Human BEC |
| 16 | 147 | 4.7 | 411 | 5 | AAE33548 Human cho |
| 17 | 147 | 4.7 | 411 | 6 | ABU03503 Angiogene |
| 18 | 147 | 4.7 | 411 | 7 | ADJ68589 Human hea |
| 19 | 147 | 4.7 | 411 | 8 | ADQ18725 Human sof |
| 20 | 136 | 4.3 | 474 | 3 | AAE33549 Human sec |
| 21 | 136 | 4.3 | 625 | 8 | ADL10287 Human pro |
| 22 | 136 | 4.3 | 958 | 3 | AAE33550 Human SAR |
| 23 | 136 | 4.3 | 958 | 8 | ADL83102 Human PRO |
| 24 | 134 | 4.3 | 483 | 2 | AAE33551 Mouse N-a |
| 25 | 133 | 4.2 | 484 | 2 | AAE33552 Human N-a |

| | | | | | | |
|----|-------|-----|-----|---|----------|-----------|
| 26 | 133 | 4.2 | 530 | 4 | AAE33553 | Human pro |
| 27 | 133 | 4.2 | 530 | 8 | ADQ18590 | Human sof |
| 28 | 133 | 4.2 | 531 | 5 | AAU69414 | Lung emal |
| 29 | 133 | 4.2 | 531 | 8 | ADL14283 | Human NF- |
| 30 | 128 | 4.1 | 183 | 4 | ABB68582 | Drosophil |
| 31 | 126 | 4.0 | 388 | 2 | AAE33554 | Mouse gly |
| 32 | 123.5 | 3.9 | 385 | 4 | AAE33555 | Human gly |
| 33 | 123.5 | 3.9 | 395 | 5 | ABB81554 | Human cor |
| 34 | 123.5 | 3.9 | 395 | 5 | AAE15438 | Human dru |
| 35 | 123.5 | 3.9 | 395 | 7 | ADL21086 | Novel hum |
| 36 | 122.5 | 3.9 | 395 | 8 | ADL61235 | Human tyr |
| 37 | 122.5 | 3.9 | 479 | 7 | ABM85237 | Human pro |
| 38 | 122.5 | 3.9 | 499 | 6 | ABR41139 | Human DIT |
| 39 | 120.5 | 3.8 | 479 | 2 | AAE33556 | Glycosami |
| 40 | 119.5 | 3.8 | 390 | 4 | AAE33557 | Human gly |
| 41 | 119.5 | 3.8 | 390 | 5 | ABB81556 | Human int |
| 42 | 119 | 3.8 | 481 | 7 | ABM85236 | Mouse pro |
| 43 | 118.5 | 3.8 | 418 | 5 | ABB81557 | Mouse int |
| 44 | 118 | 3.7 | 395 | 4 | AAE33558 | Mouse gly |
| 45 | 118 | 3.7 | 395 | 5 | AAU11275 | Murine in |

ALIGNMENTS

RESULT 1
AAY72641
ID AAY72641 standard; protein; 596 AA.
XX
AC AAY72641;
XX
XX
DT 02-MAY-2001 (first entry)
XX
DE Human glycosyl sulfotransferase-6 (GST-6) fragment.
XX

Human; glycosyl sulfotransferase-6; GST-6; immunosuppressive; therapy;
selectin binding inhibitor; gene therapy; inflammation;
systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adenitis;
Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
asthma; hypersensitivity; rheumatic fever; tissue rejection.

OS Homo sapiens.

XX WO200106015-A1.

XX 25-JAN-2001.

XX 19-JUL-2000; 2000WO-US019741.

XX 20-JUL-1999; 99US-0144694P.

XX 13-JUN-2000; 2000US-00593828.

XX (REGC) UNIV CALIFORNIA.

XX Rosen SD, Lee JK, Hemmerich S;

XX WPI; 2001-138471/14.

XX N-PSDB; AAD02702, AAD02703.

XX New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
diagnostic and therapeutic agent screening applications.

XX Claim 3; Fig 3; 128pp; English.

XX The present sequence is human glycosyl sulfotransferase-6 (GST-6)

CC fragment. GST is a type 2 membrane protein useful for inhibiting a
CC binding event between a selectin and a selectin ligand, which comprises
CC contacting the selectin with a non-sulphated selectin ligand, GST and a
CC small molecular agent that inhibits the sulphation activity of GST. GST

CC is also useful in inhibiting a selectin mediated binding event. GST is
CC useful in gene therapy to treat disorders such as acute or chronic
CC inflammation, systemic lupus erythematosus (SLE), rheumatoid arthritis,
CC polyarteritis nodosa, polymyositis, dermatomyositis, systemic sclerosis,
CC diabetes, glomerulonephritis, myasthenia gravis, Sjogren's syndrome,
CC Hashimoto's disease, Grave's disease, adrenalitis, hypoparathyroidism,
CC pernicious anaemia, demyelinating diseases, cirrhosis, ulcerative
CC colitis, dermatitis, myocarditis, regional enteritis, adult respiratory
CC distress syndrome, infantile eczema, psoriasis lichen planus, allergic
CC rhinitis, bronchial asthma, hypersensitivity, rheumatic fever and tissue
CC rejection during transplantation
XX
XX Sequence 596 AA;

Query Match 100.0%; Score 3147; DB 4; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.6e-307;
Matches 596; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MNRNGAMDMVDVDAHYKMFDFDHGNSPMASIQEAEQAAEFKKRWTFQVNVTFQMESTIT 60
DB 1 MNRNGAMDMVDVDAHYKMFDFDHGNSPMASIQEAEQAAEFKKRWTFQVNVTFQMESTIT 60
QY 61 RIAYVFGPVYINVSSCRFIDSSNPGQLQISLVNNTHEHVSIVTDYHNLKTRFNYLFGGPF 120
DB 61 RIAYVFGPVYINVSSCRFIDSSNPGQLQISLVNNTHEHVSIVTDYHNLKTRFNYLFGGPF 120
QY 121 ASVADQGITRFGGLGTQAIKVPVRHDIRIIFPFGKFNIAVGLILCISLVILTFQWRFYLS 180
DB 121 ASVADQGITRFGGLGTQAIKVPVRHDIRIIFPFGKFNIAVGLILCISLVILTFQWRFYLS 180
QY 181 FRKLMRWILILVIALWFIELLDVWSTCSQPICAKWTREAGSKSLSEGGHMDLPDVV 240
DB 181 FRKLMRWILILVIALWFIELLDVWSTCSQPICAKWTREAGSKSLSEGGHMDLPDVV 240
QY 241 ITSLPGSGAAILKQLFPFNSSDFLYIRVPTAYIDIPETELIDSFVDACEWKVSDIRSGHF 300
DB 241 ITSLPGSGAAILKQLFPFNSSDFLYIRVPTAYIDIPETELIDSFVDACEWKVSDIRSGHF 300
QY 301 RLLRGWLQSLVQDTKLQIHILHEPNRGKLAQYFAMNKKQKFKRRESLPEQSQMKG 360
DB 301 RLLRGWLQSLVQDTKLQIHILHEPNRGKLAQYFAMNKKQKFKRRESLPEQSQMKG 360
QY 361 AFRDABYIRALRRHLVYVSARPVLSLSSGSWTLKLFHFFQEVLGASMRALYIVRDPRAW 420
DB 361 AFRDABYIRALRRHLVYVSARPVLSLSSGSWTLKLFHFFQEVLGASMRALYIVRDPRAW 420
QY 421 IYSMLYNSKPSLYSLKNVPEHLAKLFKIEGGKGNLNSGYAFEPYELRKLSEKSKSNAV 480
DB 421 IYSMLYNSKPSLYSLKNVPEHLAKLFKIEGGKGNLNSGYAFEPYELRKLSEKSKSNAV 480
QY 481 SLISHLWLANTAAALRINTDPTSYQLVKFEDIVHPQKTTERRIFAFGLIPISLASLNQ 540
DB 481 SLISHLWLANTAAALRINTDPTSYQLVKFEDIVHPQKTTERRIFAFGLIPISLASLNQ 540
QY 541 ILFATSNLFLVYPYEGEISPTNTNVKQNLPRDEIKLIENICWTLMDRLGYPKFMD 596
DB 541 ILFATSNLFLVYPYEGEISPTNTNVKQNLPRDEIKLIENICWTLMDRLGYPKFMD 596

RESULT 2
AAY72642
ID AAY72642 standard; protein; 1222 AA.
XX
AC AAY72642;
XX
DT 02-MAY-2001 (first entry)
XX
DE Human glycosyl sulfotransferase-6 (GST-6).

XX Human; glycosyl sulfotransferase-6; GST-6; immunosuppressive; therapy;
KW selectin binding inhibitor; gene therapy; inflammation;
KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
KW polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;

KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenalitis;
KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
XX asthma; hypersensitivity; rheumatic fever; tissue rejection.
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH Domain 851..1223
FT /label= C-terminal_sulfotransferase_domain
XX
XX WO200106015-A1.
XX
XX 25-JAN-2001.
XX
XX 19-JUL-2000; 2000WO-US019741.
XX
XX 20-JUL-1999; 99US-0144694P.
PR 13-JUN-2000; 2000US-00593828.
XX
XX (REGC) UNIV CALIFORNIA.
PA
XX Rosen SD, Lee JK, Hemmerich S;
PI WPI: 2001-138471/14.
XX N-PSDB; AAD02702, AAD02704.
DR
XX
XX New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
PT diagnostic and therapeutic agent screening applications.
XX
XX Claim 3; Fig 5B; 128pp; English.
XX
CC The present sequence is human glycosyl sulfotransferase-6 (GST-6). GST is
CC a type 2 membrane protein useful for inhibiting a binding event between a
CC selectin and a selectin ligand, which comprises contacting the selectin
CC with a non-sulphated selectin ligand, GST and a small molecular agent
CC that inhibits the sulphation activity of GST. GST is also useful in
CC inhibiting a selectin mediated binding event. GST is useful in gene
CC therapy to treat disorders such as acute or chronic inflammation,
CC systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis
CC nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
CC disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious
CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis, distress
CC dermatitis, myocarditis, regional enteritis, adult respiratory distress
CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
CC during transplantation
XX
SQ Sequence 1222 AA;

Query Match 100.0%; Score 3147; DB 4; Length 1222;
Best Local Similarity 100.0%; Pred. No. 4.9e-307;
Matches 596; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNRNGAMDMVDVDAHYKMFDFDHGNSPMASIQEAEQAAEFKKRWTFQVNVTFQMESTIT 60
DB 627 MNRNGAMDMVDVDAHYKMFDFDHGNSPMASIQEAEQAAEFKKRWTFQVNVTFQMESTIT 686
QY 61 RIAYVFGPVYINVSSCRFIDSSNPGQLQISLVNNTHEHVSIVTDYHNLKTRFNYLFGGPF 120
DB 687 RIAYVFGPVYINVSSCRFIDSSNPGQLQISLVNNTHEHVSIVTDYHNLKTRFNYLFGGPF 746
QY 121 ASVADQGITRFGGLGTQAIKVPVRHDIRIIFPFGKFNIAVGLILCISLVILTFQWRFYLS 180
DB 747 ASVADQGITRFGGLGTQAIKVPVRHDIRIIFPFGKFNIAVGLILCISLVILTFQWRFYLS 806
QY 181 FRKLMRWILILVIALWFIELLDVWSTCSQPICAKWTREAGSKSLSEGGHMDLPDVV 240
DB 807 FRKLMRWILILVIALWFIELLDVWSTCSQPICAKWTREAGSKSLSEGGHMDLPDVV 866
QY 241 ITSLPGSGAAILKQLFPFNSSDFLYIRVPTAYIDIPETELIDSFVDACEWKVSDIRSGHF 300

||||| 867 ITSLPGSABILKQLFNSSDFLYIRVPTAYIDIPETELEIDSVDACEWKVSDIRSGHF 926
QY 301 RLLRGWLSQSLVODTKLHLQNIHLHPEPNRGKLAQYFAMNKKRKFKRRESLPEORSQWKG 360
Db 927 RLLRGWLSQSLVODTKLHLQNIHLHPEPNRGKLAQYFAMNKKRKFKRRESLPEORSQWKG 986
QY 361 AFDRDAEYIRALRRHLVYPSARPVLSSSGSWTLKLHFFQEVLGASMRALYIYVRDPRAW 420
Db 987 AFDRDAEYIRALRRHLVYPSARPVLSSSGSWTLKLHFFQEVLGASMRALYIYVRDPRAW 1046
QY 421 IYMLYNSKPSLYSLKNVPEHLAKLFKEGKGKCNLNSGYAFYEPLRKELSKSKSNV 480
Db 1047 IYMLYNSKPSLYSLKNVPEHLAKLFKEGKGKCNLNSGYAFYEPLRKELSKSKSNV 1106
QY 481 SLLSHLWLANTAALRINTDLLPTSQYLVKPEDIVHPPQKTERIFAFGLGIPLSPASLNQ 540
Db 1107 SLLSHLWLANTAALRINTDLLPTSQYLVKPEDIVHPPQKTERIFAFGLGIPLSPASLNQ 1166
QY 541 ILFATSNLFLPYEGEISPTNTNWQNLPRDEIKLIENICWTLMDRLGYPKFM 596
Db 1167 ILFATSNLFLPYEGEISPTNTNWQNLPRDEIKLIENICWTLMDRLGYPKFM 1222

RESULT 3
AAE33541
ID AAE33541 standard; protein; 1212 AA.
XX AAE33541;
AC AAE33541;
DT 16-APR-2003 (first entry)
DE Human novel CpG-associated gene 1 (NCAG1) encoded protein #1.
KW Human; novel CpG-associated gene 1; bipolar disorder; neuroprotective;
KW NCAG1; mood disorder; chromosome 18.
XX Homo sapiens.
XX WO2002101044-A2.
XX 19-DEC-2002.
XX 06-JUN-2002; 2002WO-EP006316.
XX 11-JUN-2001; 2001EP-00202214.
XX (JANC) JANSSEN PHARM NV.
XX Del-Pavero JPL, Van Broeckhoven C;
XX WPI; 2003-148807/14.
XX N-PSDB; AAD50032.
XX New brain expressed genes (designated novel CpG-associated gene 1
PT (NCAG1)) and its encoded protein, useful as diagnostic markers for
PT bipolar or mood disorders, and as targets for developing drugs for the
PT treatment these disorders.
XX Claim 14; Col 38-42; 26pp; English.
XX The invention relates to novel CpG-associated gene 1 (NCAG1) brain-
CC expressed gene and its encoded protein. The NCAG1 nucleic acid or its
CC encoded protein is useful as a diagnostic marker for bipolar disorder
CC such as mood disorders. They are also useful as targets for developing
CC drugs, as well as for target validation, for the treatment of bipolar
CC disorders. The present sequence is human NCAG1 protein. NCAG1 gene is
CC located at chromosome 18
XX Sequence 1212 AA;

Query Match 99.8%; Score 3142; DB 6; Length 1212;
Best Local Similarity 99.8%; Pred. No. 1.5e-306;

Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MNRNGAMMDVMDAHYKMFDFHGHGNSPMASIOEAEQAABFKKRWTFQVNVVTFQMESTIT 60
Db 617 MNRNGAMMDVMDAHYKMFDFHGHGNSPMASIOEAEQAABFKKRWTFQVNVVTFQMESTIT 676
QY 61 RIAYVFVGPYINVSRCRFDSSNPGLQISLVNNTTEHVSVITVDYHNLKTRFNYLGGGPF 120
Db 677 RIAYVFVGPYINVSRCRFDSSNPGLQISLVNNTTEHVSVITVDYHNLKTRFNYLGGGPF 736
QY 121 ASVADQOITRFGLGTOAIVKPVHRDRIIPFGFKFNIAVGLILICISLVILTFQWRFYLS 180
Db 737 ASVADQOITRFGLGTOAIVKPVHRDRIIPFGFKFNIAVGLILICISLVILTFQWRFYLS 796
QY 181 FRKLMRWILILVIALWFIELLDVMTSCSQPICAKWTRTEAGSKSLSSSEGHMDLPDV 240
Db 797 FRKLMRWILILVIALWFIELLDVMTSCSQPICAKWTRTEAGSKSLSSSEGHMDLPDV 856
QY 241 ITSLPGSABILKQLFNSSDFLYIRVPTAYIDIPETELEIDSVDACEWKVSDIRSGHF 300
Db 857 ITSLPGSABILKQLFNSSDFLYIRVPTAYIDIPETELEIDSVDACEWKVSDIRSGHF 916
QY 301 RLLRGWLSQSLVODTKLHLQNIHLHPEPNRGKLAQYFAMNKKRKFKRRESLPEORSQWKG 360
Db 917 RLLRGWLSQSLVODTKLHLQNIHLHPEPNRGKLAQYFAMNKKRKFKRRESLPEORSQWKG 976
QY 361 AFDRDAEYIRALRRHLVYPSARPVLSSSGSWTLKLHFFQEVLGASMRALYIYVRDPRAW 420
Db 977 AFDRDAEYIRALRRHLVYPSARPVLSSSGSWTLKLHFFQEVLGASMRALYIYVRDPRAW 1036
QY 421 IYMLYNSKPSLYSLKNVPEHLAKLFKEGKGKCNLNSGYAFYEPLRKELSKSKSNV 480
Db 1037 IYMLYNSKPSLYSLKNVPEHLAKLFKEGKGKCNLNSGYAFYEPLRKELSKSKSNV 1096
QY 481 SLLSHLWLANTAALRINTDLLPTSQYLVKPEDIVHPPQKTERIFAFGLGIPLSPASLNQ 540
Db 1097 SLLSHLWLANTAALRINTDLLPTSQYLVKPEDIVHPPQKTERIFAFGLGIPLSPASLNQ 1156
QY 541 ILFATSNLFLPYEGEISPTNTNWQNLPRDEIKLIENICWTLMDRLGYPKFM 596
Db 1157 ILFATSNLFLPYEGEISPTNTNWQNLPRDEIKLIENICWTLMDRLGYPKFM 1212

RESULT 4
ABU11849
ID ABU11849 standard; protein; 1222 AA.
XX ABU11849;
AC ABU11849;
DT 12-FEB-2003 (first entry)
DE Human secreted protein SECP-4, INCYTE 3441255CD1.
KW Human; SECP; secreted protein; micro-array; liver disease; hepatitis;
KW cirrhosis; cell proliferative disease; cancer; atherosclerosis;
KW neurological disorder; epilepsy; Huntington's disease; stroke;
KW cardiovascular disorder; hypertension; angina pectoris; allergy;
KW myocardial infarction; immune disorder; inflammatory disorder; AIDS;
KW hypothyroidism; acquired immunodeficiency syndrome; Cushing's syndrome;
KW developmental disorder.
XX Homo sapiens.
XX WO200286069-A2.
XX 31-OCT-2002.
XX 19-APR-2002; 2002WO-US012464.
XX 20-APR-2001; 2001US-0285207P.
PR 27-APR-2001; 2001US-0287114P.
PR 03-MAY-2001; 2001US-0288640P.
PR 11-MAY-2001; 2001US-0290516P.

18-MAY-2001; 2001US-0292184P.
 21-DEC-2001; 2001US-0343553P.
 13-FEB-2002; 2002US-0357002P.
 20-FEB-2002; 2002US-0358279P.
 19-MAR-2002; 2002US-0366041P.
 (INCY-) INCYTE GENOMICS INC.
 Klammer AA, Hafalia AJA, Duggan BM, Warren BA, Emerling BM;
 Tribouley CM, Arvizu CS, Ronchelli CD, Nguyen DB, Kallik DA, Yue H;
 Au-Young JK, Ramkumar J, Li JX, Thangavelu K, Gietzen KJ, Ding L;
 Baughn MR, Yao MG, Wallia NK, Mason PM, Lal PG, Gaul RC, Reddy R;
 Becha SD, Sapperstein SK, Richardson TW, Tran UK, Elliott VS;
 Tang YT, Azimzai Y, Yan L, Xu Y;
 WPI; 2003-093118/08.
 N-PSDB; ABX49951.
 New human secreted proteins (SECP) useful for diagnosing, treating and
 preventing diseases or conditions associated with the aberrant SECP
 expression e.g. cancer, AIDS, atherosclerosis, epilepsy, allergies,
 hepatitis, cirrhosis.
 Claim 1; Page 137-140; 192pp; English.
 The invention relates to an isolated polypeptide comprising any of 30
 secreted human proteins (SECP1-SECP30) appearing as ABU1846-REU1875, a
 naturally occurring amino acid sequence at least 90-98 % identical to the
 sequences, or a biologically active or immunogenic fragment of the
 polypeptide. Also included are an isolated polynucleotide encoding SECP,
 (including a polynucleotide sequence at least 90-98 % identical to the
 sequences, their complements, RNA equivalents or fragments comprising at
 least 60 contiguous nucleotides) a recombinant polynucleotide comprising
 a promoter sequence operably linked to the SECP polynucleotide, a cell
 transformed with the recombinant polynucleotide, a transgenic organism
 comprising the recombinant polynucleotide, an anti-SECP antibody,
 screening for ant/agonists of SECP, generating an expression profile of a
 sample containing the polynucleotides and an array comprising different
 nucleotide molecules affixed at distinct physical locations on a solid
 substrate, where at least one nucleotide molecule comprises a first
 oligonucleotide or polynucleotide sequence specifically hybridisable with
 at least 30 contiguous nucleotides of the target polynucleotide. The
 polypeptides and polynucleotides are useful in diagnosing, treating and
 preventing diseases or conditions associated with the decreased
 expression or overexpression of SECP, such as liver (e.g. hepatitis,
 cirrhosis), cell proliferative (e.g. cancer, atherosclerosis),
 neurological (e.g. epilepsy, Huntington's disease, stroke),
 cardiovascular (e.g. hypertension, angina pectoris, myocardial
 infarction), immune/inflammatory (e.g. acquired immunodeficiency syndrome
 (AIDS), allergies) and developmental (e.g. hypothyroidism, Cushing's
 syndrome) disorders (many other diseases and conditions are given in the
 specification). These are also useful in assessing the effects of
 exogenous compounds on the expression of nucleic acid and amino acid
 sequences of SECP. The SECP or its fragments are useful in screening
 compounds for effectiveness as agonist or antagonist of the polypeptides,
 or in altering the expression of the target polynucleotide and compounds
 that specifically bind to or modulate the activity of the polypeptide.
 The micro-array is useful in monitoring or measuring protein-protein
 interactions, drug-target interactions, and gene expression profiles. The
 present sequence is a SECP protein of the invention

687 RIAYVFGPYINVSRCRFDSSNPGLOISLVNNTTEHVSVITVDYHNLKTRFNLYGFGGF 746
 121 ASVADQGOITRGLCTQAIKVPVRHDIIFPPGPKFNIAVGLICISLVILFQWRFYLS 180
 747 ASVADQGOITRGLCTQAIKVPVRHDIIFPPGPKFNIAVGLICISLVILFQWRFYLS 806
 181 FRKLNRWILILVIALWFIELLDVWSTCSQPICAKWTTRTEAGSKSKLSSEGHHMDLPDVV 240
 807 FRKLNRWILILVIALWFIELLDVWSTCSQPICAKWTTRTEAGSKSKLSSEGHHMDLPDVV 866
 241 ITSLPGSGAEILKQLFFNSDFLYIRVPTAYIDIPETELEIDSFVDACSWKVSDIRSGHF 300
 867 ITSLPGSGAEILKQLFFNSDFLYIRVPTAYIDIPETELEIDSFVDACSWKVSDIRSGHF 926
 301 RLIRGLWSLIVQDTKLHLQNIHLHFNRCGLAQYFAMNKDKKPKRRRESLPEQRSOMKG 360
 927 RLIRGLWSLIVQDTKLHLQNIHLHFNRCGLAQYFAMNKDKKPKRRRESLPEQRSOMKG 986
 361 AFDRDAEYIRALRRHLVYVPSARPVLSSSSGSWTLLKHFQEVILGASMRALYIVRDPRAW 420
 987 AFDRDAEYIRALRRHLVYVPSARPVLSSSSGSWTLLKHFQEVILGASMRALYIVRDPRAW 1046
 421 IYSMLYNSKPSLYSLKQVPEHLAKLFKIEGKGKCNLSGYAFEPYELPKELSKSKSNV 480
 1047 IYSMLYNSKPSLYSLKQVPEHLAKLFKIEGKGKCNLSGYAFEPYELPKELSKSKSNV 1106
 481 SLSSHLLWLTAAARINTDLLPTSQVLVKFEDIHVFPOKTTTERIFAFILGIPLSPASLNO 540
 1107 SLSSHLLWLTAAARINTDLLPTSQVLVKFEDIHVFPOKTTTERIFAFILGIPLSPASLNO 1166
 541 ILPATSTNLFPYLPYEGEISPTNTNVWKNQLPRDEIKLIENICWTLMDRLGYPKFM 596
 1167 ILPATSTNLFPYLPYEGEISPTNTNVWKNQLPRDEIKLIENICWTLMDRLGYPKFM 1222

RESULT 5
 AAY72643
 ID AAY72643 standard; protein; 1207 AA.
 XX AAY72643;
 AC AAY72643;
 XX
 DT 02-MAY-2001 (first entry)
 XX
 DE Mouse glycosyl sulfotransferase-6 (GST-6).
 XX
 KW Mouse; glycosyl sulfotransferase-6; GST-6; immunosuppressive; therapy;
 selectin binding inhibitor; gene therapy; inflammation;
 KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
 KW polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
 KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; anaemia;
 KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
 KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
 KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
 KW asthma; hypersensitivity; rheumatic fever; tissue rejection.
 XX
 OS Mus sp.
 XX
 PN WO200106015-A1.
 XX
 PD 25-JAN-2001.
 XX
 XX 19-JUL-2000; 2000WO-US019741.
 XX
 PR 20-JUL-1999; 99US-0144694P.
 PR 13-JUN-2000; 2000US-00593828.
 XX
 PA (REGC) UNIV CALIFORNIA.
 XX
 PI Rosen SD, Lee JK, Hemmerich S;
 XX
 DR WPI; 2001-138471/14.
 DR N-PSDB; AAD02705, AAD02706.
 XX

Query Match 99.8%; Score 3142; DB 6; Length 1222;
 Best Local Similarity 99.8%; Pred. No. 1.6e-306;
 Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 1 MNRVNGAMDVWDHYKWFDFHNGNSPMASIQAEQADEFKRWTFVNVTFQWSTIT 60
 627 MNRVNGAMDVWDHYKWFDFHNGNSPMASIQAEQADEFKRWTFVNVTFQWSTIT 686
 61 RIAYVFGPYINVSRCRFDSSNPGLOISLVNNTTEHVSVITVDYHNLKTRFNLYGFGGF 120

PT New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
 XX diagnostic and therapeutic agent screening applications.

PS Example 2; Fig 6B; 128pp; English.

XX The present sequence is mouse glycosyl sulfotransferase-6 (GST-6). GST is
 CC a type 2 membrane protein useful for inhibiting a binding event between a
 CC selectin and a selectin ligand, which comprises contacting the selectin
 CC with a non-sulphated selectin ligand, GST and a small molecular agent
 CC that inhibits the sulphation activity of GST. GST is also useful in
 CC inhibiting a selectin mediated binding event. GST is useful in gene
 CC therapy to treat disorders such as acute or chronic inflammation,
 CC systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis
 CC nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
 CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
 CC disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious
 CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
 CC dermatitis, myocarditis, regional enteritis, adult respiratory distress
 CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
 CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
 CC during transplantation

XX Sequence 1207 AA;

Query Match 90.1%; Score 2835; DB 4; Length 1207;
 Best Local Similarity 88.6%; Pred. No. 1.4e-275;
 Matches 528; Conservative 36; Mismatches 32; Indels 0; Gaps 0;

Qy 1 MNRNGAMVDWDAHYKQWFDHGHNSPMASIQBAEQAAEFKKRWTFQVNVTFHMESTIT 60
 Db 612 MNRNGAMVDWDAHYKQWFDHGHNSPMASIQBAEQAAEFKKRWTFQVNVTFHMESTIT 671

Qy 61 RIAYVFGPYNVSSCFIDSSNPGLOISLVNNTHEHVSIVTDYHNLKTRFNLGPGGF 120
 Db 672 RIAYVFGPYNVSSCFIDSSNPGLOISLVNNTHEHVSIVTDYHNLKTRFNLGPGGF 731

Qy 121 ASVADQOQITRFGLTQAIKVPVRHRIIPFGFKFNIAGVGLIICISILVITFQWRFYLS 180
 Db 732 ASVANQOQITRFGLTQAIKVPVRHRIIPFGFKFNIAGVGLIICISILVITFQWRFYLS 791

Qy 181 FRKLMRWILVIALWFIELLDVNSTCQPICAKWTREAGSKSLSEGHMDLPDVV 240
 Db 792 FRKLMRWILVIALWFIELLDVNSTCQPICAKWTREAGSKSLSEGHMDLPDVV 851

Qy 421 IYSMLYNSKPSLYSLKNVPEHLAKLFTKEGKGNLNSGYAFEPYELPKLSKSNV 480
 Db 1032 IYSVLYGSKPSLYSLKNVPEHLAKLFTKEGKGNLNSGYAFEPYELPKLSKSNV 1091

Qy 481 SLLSHLWANTAAALRINTDLLPTSYQLVKFEDIVHPQKTERIFAPLGIPLSPASLNQ 540
 Db 1092 SLLSHLWANTAAALRINTDLLPTSYQLVKFEDIVHPQKTERIFAPLGIPLSPASLNQ 1151

Qy 541 ILPATSTNLFYLPVEGEISPTNTNVWKNLPRDEIKLIENICWTLMRLDLYGPKFMD 596
 Db 1152 MLFATSTNLFYLPVEGEISPTNTNVWKNLPRDEIKLIENICWTLMRLDLYGPKFMD 1207

RESULT 6
 AAE33542
 ID AAE33542 standard; protein; 1207 AA.
 XX

AC AAE33542;
 XX 16-APR-2003 (first entry)
 DT
 XX Human novel Cpg-associated gene 1 (NCAG1) encoded protein #2.
 DE
 XX Human; novel Cpg-associated gene 1; bipolar disorder; neuroprotective;
 KW NCAG1; mood disorder; chromosome 18.
 KW
 XX Homo sapiens.
 OS
 XX WO2002101044-A2.
 PN
 XX 19-DEC-2002.
 PD
 XX 06-JUN-2002; 2002WO-EP006316.
 PF
 XX 11-JUN-2001; 2001EP-00202214.
 PR
 XX (JANC) JANSSEN PHARM NV.
 PA
 XX Del-Pavero JPL, Van Broeckhoven C;
 PI
 XX WPI; 2003-148807/14.
 DR
 XX N-PSDB; AAD50033.
 DR
 XX New brain expressed genes (designated novel Cpg-associated Gene 1
 PT (NCAG1)) and its encoded protein, useful as diagnostic markers for
 PT bipolar or mood disorders, and as targets for developing drugs for the
 PT treatment of these disorders.
 PT
 XX Claim 16; Col 48-51; 26pp; English.
 PS
 XX The invention relates to novel Cpg-associated gene 1 (NCAG1) brain-
 CC expressed gene and its encoded protein. The NCAG1 nucleic acid or its
 CC encoded protein is useful as a diagnostic marker for bipolar disorder
 CC such as mood disorders. They are also useful as targets for developing
 CC drugs, as well as for target validation, for the treatment of bipolar
 CC disorders. The present sequence is human NCAG1 protein. NCAG1 gene is
 CC located at chromosome 18
 CC
 XX Sequence 1207 AA;

Query Match 90.1%; Score 2835; DB 6; Length 1207;
 Best Local Similarity 88.6%; Pred. No. 1.4e-275;
 Matches 528; Conservative 36; Mismatches 32; Indels 0; Gaps 0;

Qy 1 MNRNGAMVDWDAHYKQWFDHGHNSPMASIQBAEQAAEFKKRWTFQVNVTFHMESTIT 60
 Db 612 MNRNGAMVDWDAHYKQWFDHGHNSPMASIQBAEQAAEFKKRWTFQVNVTFHMESTIT 671

Qy 61 RIAYVFGPYNVSSCFIDSSNPGLOISLVNNTHEHVSIVTDYHNLKTRFNLGPGGF 120
 Db 672 RIAYVFGPYNVSSCFIDSSNPGLOISLVNNTHEHVSIVTDYHNLKTRFNLGPGGF 731

Qy 121 ASVADQOQITRFGLTQAIKVPVRHRIIPFGFKFNIAGVGLIICISILVITFQWRFYLS 180
 Db 732 ASVANQOQITRFGLTQAIKVPVRHRIIPFGFKFNIAGVGLIICISILVITFQWRFYLS 791

Qy 181 FRKLMRWILVIALWFIELLDVNSTCQPICAKWTREAGSKSLSEGHMDLPDVV 240
 Db 792 FRKLMRWILVIALWFIELLDVNSTCQPICAKWTREAGSKSLSEGHMDLPDVV 851

Qy 241 ITSPLGSGAEILKQLFFNSDPLIYRPTAYIDIPETELETDSFVDAEAKWVSIRSGHF 300
 Db 852 ITSPLGSGAEILKQLFFNSDPLIYRPTAYIDIPETELETDSFVDAEAKWVSIRSGHF 911

Qy 301 RLLRGWLSQVQDTKHLQNIHLHPEHNRKLAQYFAMNKKRKFRRRESLPEQRSQMKG 360
 Db 912 HLLRGWLSQVQDTKHLQNIHLHPEHNRKLAQYFAMNKKRKFRRRESLPEQRSQMKG 971

Qy 361 AFDRDAEYIRALRRHLVYYPSPARVLSLSSGWSLTKLHFFQEVIGASNRALYIVRDPRAW 420
 Db 972 AFDRDAEYIRALRRHLVYYPSPARVLSLSSGWSLTKLHFFQEVIGASNRALYIVRDPRAW 1031

Qy 421 IYSMLYNSKPSLYSLKNVPEHLAKLFTKEGKGNLNSGYAFEPYELPKLSKSNV 480
 Db 1032 IYSVLYGSKPSLYSLKNVPEHLAKLFTKEGKGNLNSGYAFEPYELPKLSKSNV 1091

Qy 481 SLLSHLWANTAAALRINTDLLPTSYQLVKFEDIVHPQKTERIFAPLGIPLSPASLNQ 540
 Db 1092 SLLSHLWANTAAALRINTDLLPTSYQLVKFEDIVHPQKTERIFAPLGIPLSPASLNQ 1151

Qy 541 ILPATSTNLFYLPVEGEISPTNTNVWKNLPRDEIKLIENICWTLMRLDLYGPKFMD 596
 Db 1152 MLFATSTNLFYLPVEGEISPTNTNVWKNLPRDEIKLIENICWTLMRLDLYGPKFMD 1207

Db 972 PFORDAEYIRALRRHLVYYPARPVLSLSSGSWTLKLHFFQEVGLTSMRALYIVRDPRAW 1031
QY 421 IYSLMLYNSKPSLYSLKNVPBHLAKLPKIEGKGKCNLNSGYAFEPPLRKLKSKSNAV 480
Db 1032 IYSLYSGKPSLYSLKNVPBHLAKLPKIEGKGKCNLSNGYAFESLKKELISQSNAI 1091
QY 481 SLASHLWLANTAARLINTDLPSTYOLVKFEDIVHFPQKTTERRIFAFGLGIPLSASLNQ 540
Db 1092 SLASHLWVANTAARLINTDLPYHLVKFEDIVHFPQKTTERRIFAFGLGIPLSASLNQ 1151
QY 541 ILFATSTNLYLPYEGEISPTNTNVKQNLPRDEIKLIENICWTLMDRLGYPKFMD 596
Db 1152 MLFATSTNLYLPYEGEISPTNTNIWKTNLPRDEIKLIENICWTLMDHLGYPKFMD 1207

RESULT 7
AAB93735
ID AAB93735 standard; protein; 755 AA.
XX AC AAB93735;
XX 26-JUN-2001 (first entry)
XX Human protein sequence SEQ ID NO:13382.
DE Human; primer; detection; diagnosis; antisense therapy; gene therapy.
KW Homo sapiens.
OS EP1074617-A2.
PN 07-FEB-2001.
PD 28-JUL-2000; 2000EP-00116126.
PF 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX (HELI-) HELIX RES INST.
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiya T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
DR Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.
XX Claim 8; SEQ ID NO 13382; 2537pp + Sequence Listing; English.
XX The present invention describes primer sets for synthesising 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dr primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and

CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent CC oligonucleotides, all of which are used in the exemplification of the CC present invention
XX Sequence 755 AA;
SQ Query Match 57.9%; Score 1821; DB 4; Length 755;
Best Local Similarity 99.4%; Pred. No. 1.le-173;
Matches 342; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 MNRNGAMMDVMDAHYKMFWDHGHNSPMASIQEAEQAAEFKKRWTFQVNVTFQMESTIT 60
Db 411 MNRNGAMMDVMDAHYKMFWDHGHNSPMASIQEAEQAAEFKKRWTFQVNVTFQMESTIT 470
QY 61 RIAYFYGPYINVSRCRFDSSNPGLOISLVNNTNTEHVSIIVTDYHNLKTRNLYLGGGF 120
Db 471 RIAYFYGPYINVSRCRFDSSNPGLOISLVNNTNTEHVSIIVTDYHNLKTRNLYLGGGF 530
QY 121 ASVADQGQITRFGLGTOAIVKPVHRDRIIFPFGFKFNIAVGLILCISLVILTFQWRFYLS 180
Db 531 ASVADQGQITRFGLGTOAIVKPVHRDRIIFPFGFKFNIAVGLILCISLVILTFQWRFYLS 590
QY 181 FRKLNRWILILVIALWFIELLDVWSTCSQPICAKWTTRTEAGSKKSLSEGHMDLPDVV 240
Db 591 FRKLNRWILILVIALWFIELLDVWSTCSQPICAKWTTRTEAGSKKSLSEGHMDLPDVV 650
QY 241 ITSLSGSGAEILKQLPFNSSDFLYIRVPTAYIDIPETELEDISFVDACEWKVSDIRSGHF 300
Db 651 ITSLSGSGAEILKQLPFNSSDFLYIRVPTAYIDIPETELEDISFVDACEWKVSDIRSGHF 710
QY 301 RLLRGWLQSLVQDTKLHLQNIHLHFNPRGKLAQYFAMNKDKKK 344
Db 711 RLLRGWLQSLVQDTKLHLQNIHLHFNPRGKLAQYFAMNKDKKK 754

RESULT 8
ADP56673
ID ADP56673 standard; protein; 755 AA.
XX AC ADP56673;
XX 12-AUG-2004 (first entry)
DT Human protein which is a chondroitin 6 sulphotransferase 3 homologue.
DE Human protein which is a chondroitin 6 sulphotransferase 3 homologue.
XX Cancer detection; large intestine; oesophagus; stomach; lungs; pancreas;
KW liver; kidney; colon; human; chondroitin 6 sulphotransferase 3; CH6T3.
XX Homo sapiens.
XX JP2004147505-A.
XX 27-MAY-2004.
XX 28-OCT-2002; 2002JP-00312927.
XX 28-OCT-2002; 2002JP-00312927.
XX (DOKU-) DOKURITSU GYOSEI HOJIN SANGYO GIJUTSU SO.
PA (FJRE) FUJIREBIO KK.
PA (SEK) SEIKAGAKU KOGYO CO LTD.
XX WPI; 2004-434537/41.
DR N-PSDB; ADP56672.
XX Novel nucleic acid such as DNA, useful as probe for identifying or
PT detecting cancerous tissue of esophagus, stomach, lungs, pancreas, liver,
PT kidney or colon, preferably large intestine tissue.
XX Example 1; SEQ ID NO 2; 24pp; Japanese.
XX The invention relates to a novel nucleic acid having 40-1000 base pairs
CC

CC and comprising a sequence which is complementary to 41-2308 nucleotides
 CC of a fully defined sequence of 3288 base pairs as given in the
 CC specification. The polynucleotide of the invention may be useful for
 CC rapid and highly reliable detection of cancerous tissue derived from the
 CC large intestine, oesophagus, stomach, lungs, pancreas, liver, kidney or
 CC colon. The current sequence is that of the human protein of the invention
 CC which is a chondroitin 6 sulphotransferase 3 (CHST3) homologue.
 XX
 SQ Sequence 755 AA;

Query Match 57.9%; Score 1821; DB 8; Length 755;
 Best Local Similarity 99.4%; Pred. No. 1.1e-173; Mismatches 1; Indels 0; Gaps 0;
 Matches 342; Conservative 1;

QY 1 MNRVNGAMDMVDVDAHYKMFDFDHGNSPMASIQAEQAQAEFKKRWTFQVNVTFQMESTIT 60
 DB 411 MNRVNGAMDMVDVDAHYKMFDFDHGNSPMASIQAEQAQAEFKKRWTFQVNVTFQMESTIT 470
 QY 61 RIAYVFGPYINVSRCRFDSSNPGQLQISLVNNTVHVSIVTDYHNLKTRFNYLGFGGF 120
 DB 471 RIAYVFGPYINVSRCRFDSSNPGQLQISLVNNTVHVSIVTDYHNLKTRFNYLGFGGF 530
 QY 121 ASVADQGITRFGLTQAIKVPVRHRIIPFGFKFNIAVGLILCISLVILTFQWREYLS 180
 DB 531 ASVADQGITRFGLTQAIKVPVRHRIIPFGFKFNIAVGLILCISLVILTFQWREYLS 590
 QY 181 FRKLMRWILVIALWFIELLDVWSTCSQPTCAKWTREAGSKLSSEGHMDLDPVV 240
 DB 591 FRKLMRWILVIALWFIELLDVWSTCSQPTCAKWTREAGSKLSSEGHMDLDPVV 650
 QY 241 ITSLPGSGAEILKQFPNSSDFLYIRVPTAVIDIPETELEIDSFVDACEWKSVDIRSGHF 300
 DB 651 ITSLPGSGAEILKQFPNSSDFLYIRVPTAVIDIPETELEIDSFVDACEWKSVDIRSGHF 710
 QY 301 RLLRGWLQSLVQDTKLHLQNLHLEPNRGKLAQYFAMNKKKK 344
 DB 711 RLLRGWLQSLVQDTKLHLQNLHLEPNRGKLAQYFAMNKKKK 754

RESULT 9
 ID AAB94410 standard; protein; 480 AA.
 XX AAB94410;
 XX
 XX 26-JUN-2001 (first entry)
 DT
 DE Human protein sequence SEQ ID NO:14997.
 DE
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 KW
 OS Homo sapiens.
 XX
 XX EP1074617-A2.
 XX
 XX 07-FEB-2001.
 XX
 XX 28-JUL-2000; 2000EP-00116126.
 XX
 XX 29-JUL-1999; 99JP-00248036.
 XX
 XX 27-AUG-1999; 99JP-00300253.
 XX
 XX 11-JAN-2000; 2000JP-00118776.
 XX
 XX 02-MAY-2000; 2000JP-00183767.
 XX
 XX 09-JUN-2000; 2000JP-00241899.
 XX
 XX (HELI-) HELIX RES INST.
 XX
 XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 XX WPI; 2001-318749/34.
 XX
 XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-

PT length cDNAs defined in the specification, and for the detection and/or
 PT diagnosis of the abnormality of the proteins encoded by the full-length
 PT cDNAs.

PS Claim 8; SEQ ID NO 14997; 2537pp + Sequence Listing; English.

XX
 CC The present invention describes primer sets for synthesising 5602 full-
 CC length cDNAs defined in the specification. Where a primer set comprises:
 CC (a) an oligo-dT primer and an oligonucleotide complementary to the
 CC complementary strand of a polynucleotide which comprises one of the 5602
 CC nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the
 CC specification. The primer sets can be used in antisense therapy and in
 CC gene therapy. The primers are useful for synthesising polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAB95893
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
 CC oligonucleotides, all of which are used in the exemplification of the
 CC present invention
 XX

SQ Sequence 480 AA;

Query Match 29.2%; Score 920; DB 4; Length 480;
 Best Local Similarity 100.0%; Pred. No. 3.9e-83;
 Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNRVNGAMDMVDVDAHYKMFDFDHGNSPMASIQAEQAQAEFKKRWTFQVNVTFQMESTIT 60
 DB 308 MNRVNGAMDMVDVDAHYKMFDFDHGNSPMASIQAEQAQAEFKKRWTFQVNVTFQMESTIT 367
 QY 61 RIAYVFGPYINVSRCRFDSSNPGQLQISLVNNTVHVSIVTDYHNLKTRFNYLGFGGF 120
 DB 368 RIAYVFGPYINVSRCRFDSSNPGQLQISLVNNTVHVSIVTDYHNLKTRFNYLGFGGF 427
 QY 121 ASVADQGITRFGLTQAIKVPVRHRIIPFGFKFNIAVGLILCISLVILTF 173
 DB 428 ASVADQGITRFGLTQAIKVPVRHRIIPFGFKFNIAVGLILCISLVILTF 480

RESULT 10
 ADG27796
 ID ADG27796 standard; protein; 128 AA.
 XX
 XX ADG27796;
 XX
 XX 26-FEB-2004 (first entry)
 DT
 DE Human novel protein amino acid sequence SeqID563.
 DE
 KW antiparkinsonian; haemostatic; nootropic; neuroprotective; osteopathic;
 KW anti-HIV; protozoacide; antifungal; immunosuppressive; antirheumatic;
 KW antiarthritic; antidiabetic; anti-allergic; anti-inflammatory;
 KW anticoagulant; cytostatic; gene therapy; Parkinson's disease;
 KW Alzheimer's disease; thrombocytopaenia; osteoporosis; osteoarthritis;
 KW infection; HIV; Leishmania; malaria; fungal infection;
 KW multiple sclerosis; rheumatoid arthritis;
 KW insulin dependent diabetes mellitus; allergic reaction; food allergy;
 KW insect allergy; allergic rhinitis; haemophilia; cancer; human.
 XX
 XX Homo sapiens.
 OS
 XX WO200179254-A1.
 XX
 XX 25-OCT-2001.

XX PF 16-APR-2001; 2001WO-US008655.
 XX PR 18-APR-2000; 2000US-00552929.
 XX PR 22-SEP-2000; 2000US-00668317.
 XX PR 24-OCT-2000; 2000US-00695783.
 XX PR 01-DEC-2000; 2000US-00728628.
 XX PR 26-JAN-2001; 2001US-00770160.
 XX PR 13-FEB-2001; 2001US-00783066.
 XX PR 22-MAR-2001; 2001US-00816928.
 XX PA (HYSE-) HYSEQ INC.
 XX XX Tang YT, Asundi V, Zhou P, Xue AJ, Ren F, Zhang J, Wang J;
 PI Yang Y, Zhao QA, Goodrich RW, Liu C, Drmanac RT, Ma Y, Wang Z;
 PI Wehrman T;
 XX WPI; 2001-607699/69.
 XX XX Novel polynucleotides and encoded polypeptides (protein factors,
 PT including, e.g., cytokines, such as lymphokines, interferons, and
 PT circulating soluble factors) useful for treating, e.g., Parkinson's,
 PT Alzheimer's, HIV and cancer.
 XX XX
 PS Claim 20; SEQ ID NO 563; 153pp; English.
 XX XX
 CC This invention relates to a novel isolated DNA sequence and the mature
 CC proteins encoded by them. The invention may be useful in the development
 CC of compositions with antiparkinsonian, haemostatic, neurotropic,
 CC neuroprotective, osteoprotective, anti-HIV, protozoacide, antifungal,
 CC immunosuppressive, antiarthritis, antidiabetic, antidiabetic,
 CC antiallergic, antiinflammatory, anticoagulant or cytostatic activities.
 CC In addition, the sequences of the invention may be useful for gene
 CC therapy. The invention may be useful for the development of treatments
 CC for Parkinson's, Alzheimer's, thrombocytopaenia, osteoporosis,
 CC osteoarthritis, infections (including HIV, Leishmania, malaria, and
 CC various fungal infection), autoimmune disorders such as multiple
 CC sclerosis, rheumatoid arthritis, and insulin dependent diabetes mellitus,
 CC allergic reactions and conditions (for example food allergies, insect
 CC allergies and allergic rhinitis), coagulation disorders including
 CC haemophilia, and cancer. Note: The amino acid sequences given in table 6
 CC (SeqID 439-584) may have in frame stop codons or possible
 CC insertions/deletions as shown in the table. The sequences allocated Seq
 CC IDs 1-438 are not provided (even by reference) in the specification.
 XX XX
 SQ Sequence 128 AA;
 Query Match 20.2%; Score 637; DB 4; Length 128;
 Best Local Similarity 96.1%; Pred. No. 1.7e-55;
 Matches 122; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 470 KELS KSNVAVSLSHLWLANTAALRINTDLPTSQVLKVFEDIVHFPQKTTTIPAF 529
 Db 2 KNYQNPKNVAVSLSHLWLANTAALRINTDLPTSQVLKVFEDIVHFPQKTTTIPAF 61
 QY 530 GIPLSASLNLQILLFATSTNLFYLPYEGEISPTNTNWKNLPRDEIKLIENICWTLM 589
 Db 62 GIPLSASLNLQILLFATSTNLFYLPYEGEISPTNTNWKNLPRDEIKLIENICWTLM 121
 QY 590 GYPKFM 596
 Db 122 GYPKFM 128
 RESULT 11
 ID AAB93826
 XX AAB93826 standard; protein; 314 AA.
 XX AC AAB93826;
 XX XX 26-JUN-2001 (first entry)
 XX DE Human protein sequence SEQ ID NO:13638.

XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 XX OS Homo sapiens.
 XX PN EP1074617-A2.
 XX XX 07-FEB-2001.
 XX PF 28-JUL-2000; 2000EP-00116126.
 XX PR 29-JUL-1999; 99JP-00248036.
 XX PR 27-AUG-1999; 99JP-00300253.
 XX PR 11-JAN-2000; 2000JP-00118776.
 XX PR 02-MAY-2000; 2000JP-00183767.
 XX PR 09-JUN-2000; 2000JP-00241899.
 XX XX (HELI-) HELIX RES INST.
 XX XX Ota T, Isogai T, Nishikawa T, Hayaashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX WPI; 2001-318749/34.
 XX XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-
 PT length cDNAs defined in the specification, and for the detection and/or
 PT diagnosis of the abnormality of the proteins encoded by the full-length
 PT cDNAs.
 XX XX Claim 8; SEQ ID NO 13638; 2537pp + Sequence Listing; English.
 XX XX The present invention describes primer sets for synthesising 5602 full-
 CC length cDNAs defined in the specification, where a primer set comprises:
 CC (a) an oligo-dT primer and an oligonucleotide complementary to the 5602
 CC nucleotide sequence defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the
 CC specification. The primer sets can be used in antisense therapy and in
 CC gene therapy. The primers are useful for synthesising polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
 CC oligonucleotides, all of which are used in the exemplification of the
 CC present invention
 XX SQ Sequence 314 AA;
 Query Match 19.5%; Score 615; DB 4; Length 314;
 Best Local Similarity 100.0%; Pred. No. 1.1e-52;
 Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 224 KKSLSSEGHMHMDLPDVVITSLPGSGAEILKQLPFNSDFLYIRVPTAYIDIPETE 283
 Db 198 KKSLSSEGHMHMDLPDVVITSLPGSGAEILKQLPFNSDFLYIRVPTAYIDIPETE 257
 QY 284 FVDACEKVS DTRSGHFRLLRGWLQSLVQDTKLQNIHLHFNPRGKLAQYFAMNKD 340
 Db 258 FVDACEKVS DTRSGHFRLLRGWLQSLVQDTKLQNIHLHFNPRGKLAQYFAMNKD 314
 RESULT 12
 ID ADJ12217
 XX ADJ12217 standard; protein; 125 AA.

AC ADJ12217;
 DT 20-MAY-2004 (first entry)
 DE Human secreted protein SeqID 71.
 XX
 KW human; secreted; cancer; haematopoietic disease; anaemia;
 KW multiple myeloma; reproductive system disorder; prostatitis;
 KW inguinal hernia; musculoskeletal disease; systemic lupus erythematosus;
 KW gout; cardiovascular disease; arrhythmia; hypernatraemia; fetal disease;
 KW fetal alcohol syndrome; Down's syndrome; excretory disease;
 KW urinary incontinence; renal disorder; neural; sensory disease;
 KW Alzheimer's disease; meningitis; respiratory disease; emphysema;
 KW occupational lung disease; endocrine disease; diabetes;
 KW glomerulonephritis; digestive disease; portal hypertension;
 KW irritable bowel syndrome; epithelial disease; scleroderma;
 KW epidermolysis bullosa; cytostatic; antineoplastic; antiarrhythmic;
 KW antiasthmatic; anti-HIV; immunosuppressive; antiinflammatory;
 KW antiparkinsonian; antibacterial; osteopathic; dermatological; antitumor;
 KW immunomodulator; antiarrhythmic; cardiac; nontropic; antileptic;
 KW nephrotropic; uterine; neuroprotective; antiparkinsonian; tranquilizer;
 KW antidiabetic; anabolic; hypertensive; vulnary.
 XX
 OS Homo sapiens.
 XX
 PN US2004010132-A1.
 XX
 PD 15-JAN-2004.
 XX
 PF 30-OCT-2001; 2001US-00984429.
 XX
 PR 09-OCT-1997; 97US-0061463P.
 PR 09-OCT-1997; 97US-0061527P.
 PR 09-OCT-1997; 97US-0061529P.
 PR 09-OCT-1997; 97US-0061532P.
 PR 09-OCT-1997; 97US-0061536P.
 PR 09-OCT-1997; 97US-0071498P.
 PR 08-APR-1998; 98WO-US021142.
 PR 01-NOV-2000; 2000US-0244591P.
 XX
 PA (ROSE/) ROSEN C A.
 PA (BREW/) BREWER L A.
 PA (DUAN/) DUAN R D.
 PA (RUBEN/) RUBEN S M.
 PA (FLORENCE/) FLORENCE K A.
 PA (GREENE/) GREENE J M.
 PA (YOUNG/) YOUNG P E.
 PA (FERRIE/) FERRIE A M.
 PA (YUGG/) YU G.
 PA (FLORENCE/) FLORENCE C.
 PA (ESNER/) ESNER R.
 PA (OLSEN/) OLSEN H.
 PI Rosen CA, Brewer LA, Duan RD, Ruben SM, Florence KA, Greene JM,
 PI Young PE, Ferrie AM, Yu G, Florence C, Ebner R, Olsen H;
 XX WPI; 2004-090518/09.
 DR N-PSDB; ADJ12161.
 XX
 PT New isolated nucleic acids and polypeptides, useful for diagnosing,
 PT treating, preventing or ameliorating diseases or disorders e.g. cancer,
 PT anemia, arthritis, asthma, inflammatory bowel disease or Alzheimer's
 PT disease.
 XX
 PS Claim 11; SEQ ID NO 71; 286pp; English.
 XX
 CC This invention relates to novel polynucleotides encoding human secreted
 CC proteins. Specifically, it refers to the vectors, host cells, recombinant
 CC and synthetic methods for producing human polynucleotides, polypeptides
 CC and antibodies. Furthermore, it relates to screening methods to identify
 CC agonists and antagonists that can be used to inhibit or enhance the
 CC production and function of the secreted proteins. The present invention

CC describes these compositions as useful for diagnosing, treating or
 CC preventing disorders such as cancer, haematopoietic diseases including
 CC anaemia and multiple myeloma, reproductive system disorders including
 CC prostatitis and inguinal hernia, musculoskeletal diseases including
 CC systemic lupus erythematosus and gout, cardiovascular diseases including
 CC arrhythmia and hypernatraemia, mixed fetal diseases including fetal
 CC alcohol syndrome and Down's syndrome, excretory diseases including
 CC urinary incontinence and renal disorders, neural or sensory disease
 CC including Alzheimer's disease and meningitis, respiratory disease
 CC including emphysema and occupational lung disease, endocrine diseases
 CC including diabetes and glomerulonephritis, digestive diseases including
 CC portal hypertension and irritable bowel syndrome and connective tissue or
 CC epithelial diseases including scleroderma and epidermolysis bullosa. As
 CC such, there are various activities such as cytostatic, antineoplastic,
 CC antiarrhythmic, antiasthmatic, anti-HIV, immunosuppressive,
 CC antiinflammatory, antiparkinsonian, antibacterial, osteopathic,
 CC dermatological, antitumor, immunomodulator, antiarrhythmic, cardiac,
 CC nontropic, antileptic, nephrotropic, uterine, antiparkinsonian, tranquilizer,
 CC antiparkinsonian, tranquilizer, antidiabetic, anabolic, hypertensive and
 CC vulnary. This polypeptide is a human secreted protein of the invention.
 CC NOTE: This sequence does not appear in the printed specification but has
 CC been obtained in electronic format from the US patent office at the
 CC following web site www.seqdata.uspto.gov/sequence.html; Document ID:
 CC 20040010132.
 XX

Sequence 125 AA;

Query Match 18.0%; Score 568; DB 8; Length 125;

Best Local Similarity 100.0%; Pred. No. 1.5e-48; Mismatches 0; Indels 0; Gaps 0;

Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 MRWILILVIALWFIELLDVWSTCSQPICAKWTRTEAEGSKSLSEGHMDLPDVTSL 244
 DB 1 MRWILILVIALWFIELLDVWSTCSQPICAKWTRTEAEGSKSLSEGHMDLPDVTSL 60

QY 245 PGSGAELKQLFFNSDFLIRVPTAYIDIPETELEDIDSFVDACEWK 291

DB 61 PGSGAELKQLFFNSDFLIRVPTAYIDIPETELEDIDSFVDACEWK 107

RESULT 13

AA114415

ID AAY14415 standard; protein; 126 AA.

XX AAY14415;

AC AAY14415;

XX 17-AUG-1999 (first entry)

DT 17-AUG-1999 (first entry)

DE Human secreted protein encoded by gene 5 clone HSABG21.

XX Human; secreted protein; fusion protein; gene therapy; protein therapy;
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
 KW developmental abnormality; foetal deficiency; blood; allergy; renal;
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

OS Homo sapiens.

XX WO9919339-A1.

PN 22-APR-1999.

XX 08-OCT-1998; 98WO-US021142.

XX 09-OCT-1997; 97US-0061463P.

XX 09-OCT-1997; 97US-0061527P.

XX 09-OCT-1997; 97US-0061529P.

XX 09-OCT-1997; 97US-0061532P.

XX 09-OCT-1997; 97US-0061536P.

XX 09-OCT-1997; 97US-0071498P.

XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PA Brewer LA, Olsen HS, Duan R, Ebner R, Rosen CA, Ruben SM;
PI Florence KA, Young PE, Greene JM, Yu G, Ferrie AM, Florence C;
XX N-PSDB; AAX79015.
DR WPI; 1999-277587/23.
XX N-PSDB; AAX79015.
PT New isolated human genes and the secreted polypeptides they encode.
XX Claim 11; Page 186; 226pp; English.
XX This sequence represents a secreted human protein encoded by the gene
CC clone detailed in the descriptor line. The gene can be used to generate
CC fusion proteins by linking to the gene to a human immunoglobulin Fc
CC portion (e.g. AAX79002) for increasing the stability of the fused protein
CC as compared to the human protein only. The invention relates to 53 novel
CC genes and their fragments (nucleic acid sequences: AAX79011-X79064; amino
CC acid sequences AAY14411-Y14464) which are useful for preventing, treating
CC or ameliorating medical conditions e.g. by protein or gene therapy. Also,
CC pathological conditions can be diagnosed by determining the amount of the
CC new polypeptides in a sample or by determining the presence of mutations
CC in the new polynucleotides. Specific uses are described for each of the
CC 53 polynucleotides, based on which tissues they are most highly expressed
CC in (see AAX79011 for described uses)
XX
SQ Sequence 126 AA;

Query Match 18.0%; Score 568; DB 2; Length 126;
Best Local Similarity 100.0%; Pred. No. 1.5e-48;
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 185 MRWILLVIALWFIELDVNSTCSQPTCAKWTREARSGKSLSSGHHMDLPDVVITSL 244
DB 1 MRWILLVIALWFIELDVNSTCSQPTCAKWTREARSGKSLSSGHHMDLPDVVITSL 60
QY 245 PGSGAEILKQLFFNSDFLYIRVPTAYIDIPETELEDISFVDACEWK 291
DB 61 PGSGAEILKQLFFNSDFLYIRVPTAYIDIPETELEDISFVDACEWK 107

RESULT 14
ADN95522
ID ADN95522 standard; protein; 410 AA.
XX
AC ADN95522;
XX
DT 01-JUL-2004 (first entry)
XX
DE Human BEC/LEC-related protein sequence SeqID445.
XX
KW growth; differentiation; blood endothelial cell; BEC;
KW lymphatic endothelial cell; LEC; hereditary lymphoedema; VEGFR-3;
KW lymphatic growth agent; VEGF-C; VEGF-D; antiangiogenic; cytostatic;
KW vasotrophic; antiinflammatory; gene therapy; endothelial cell disorder;
KW inflammatory disease; cancer metastasis; lymphatic system; human.
XX
OS Homo sapiens.
XX
XX WO2003080640-A1.
XX
XX
PD 02-OCT-2003.
XX
PF 07-MAR-2003; 2003WO-US0006900.
XX
PR 07-MAR-2002; 2002US-0363019P.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
PA (LICN) LICENTIA LTD.
PI Alitalo K, Makinen T, Petrova T, Saharinen P, Saharinen J;
XX

DR WPI; 2003-876899/81.
XX N-PSDB; ADN95523.
PS Example 1; SEQ ID NO 445; 176pp; English.
XX
CC This invention relates to a method of differentially modulating the
CC growth or differentiation of blood endothelial cells (BEC) or lymphatic
CC endothelial cells (LEC) comprises contacting endothelial cells with a
CC composition comprising an agent that differentially modulates blood or
CC lymphatic endothelial cells. Treating hereditary lymphoedema comprises
CC identifying a human subject with lymphoedema and with a mutation in at
CC least one allele of a gene encoding a LEC protein, where the proviso that
CC the LEC protein is not VEGFR-3; and administering to the subject a
CC composition comprising a lymphatic growth agent selected from VEGF-C or
CC VEGF-D polypeptides and polynucleotides. The invention may be useful for
CC the development of compounds with an antiangiogenic, cytostatic,
CC vasotrophic or antiinflammatory activity or for gene therapy. The method
CC is useful in modulating the growth or differentiation of blood
CC endothelial cells or lymphatic endothelial cells, in treating hereditary
CC lymphoedema, in screening for an endothelial cell disorder or
CC predisposition to the disorder or in monitoring the efficacy or toxicity
CC of a drug on endothelial cells. The agent is useful in manufacturing a
CC medicament for the differential modulation of blood vessel endothelial
CC cell or lymphatic vessel endothelial cell growth or differentiation. The
CC treatment of hereditary lymphoedema resulting from a mutation in a
CC LEC gene or of other diseases involving the lymphatic vessels, such as
CC various inflammatory diseases and cancer metastasis via the lymphatic
CC system. The present sequence is that of a human LEC/BEC differentially
CC expressed protein which is related to the method of the invention. Note:
CC This sequence does not appear in the specification but was obtained by
CC the indexer using the source data given in table 14 of the specification.
XX
SQ Sequence 410 AA;
Query Match 4.7%; Score 147; DB 7; Length 410;
Best Local Similarity 18.2%; Pred. No. 2.6e-05;
Matches 89; Conservative 58; Mismatches 149; Indels 194; Gaps 16;
QY 187 WILVIALWFIEL-----LDVWSTC-----SQPTCAKWTREARSGKSLSS 229
DB 5 WKAVLLALASIAIOVTAIRTFKSFHTCPGLAEAGLAERLC-----ESPTFAYNLSR 59
QY 230 EGHMDLPDVVITSLPGSGAEILKQLFFNSDFLYIRVPTAYIDIPETELEDISFVDACE 289
DB 60 KTH-----ILLATRSGSSFFVQLFNOHLDFVLFEP----- 93
QY 290 WKVSDIRSGHFRLLRGWLSQVDTKLHLQNIHLHPEPNRGKLAQVPMNKDKRKFRRRE 349
DB 94 -----YHVQNTLI PRFTOG-----KSPADRRV 115
QY 350 SUPEROSQMGAFDRDAEYI-----RALRHLVYPSARV-----LS 387
DB 116 MLGASRDLLRSYDCDLYPLENIKPPVPPVNTDRIFFRGASRVLCRSREVCPPGADLV 175
QY 388 LSSGSLTKLHFFQFVLGH-----SMRALYIVRD 416
DB 176 LEEGDCVRKCGLLNLTVAABACRSHVAIKTVRVPENVNDLRALVEDPRLNLTQVLQVRD 235
QY 417 PRAWIYMLYNSKPSLYSLKQVPEHLAKLFKEGGKGCNLSNGYAFYEPRKELSKSK 476
DB 236 PGILASRSETFRDITY-----RLWRLWYGTGR-----KPNYLDVTQL-TTVCEDF 279
QY 477 SNAVS--LLSHLWLANTAALRINTDLLPTSQVLFKFDIVHFPQKTTTRIFAFGLIPL- 533
DB 280 SNSVSTGLMRPPLKLG-----KYMVRYEDLARNPMKKTETIYFGLIPLD 325
QY 534 -----SPASLNQILFATSTNLFYLPVEGEISPTNTNVKQNI-PROEIKLIENI 581
DB 326 SHVARWIONNTRGDPTLGGHKYGTVRN-----SAATAEKWRFRLSYDIVAFAPQNA 375
QY 582 CWTLMRLGCGY 591

Search completed: June 23, 2005, 08:43:25
Job time : 66.4109 secs

| | | |
|----|-----|---|
| XX | PT | Keratan sulphate 6-sulpho-transferase polypeptide - transfers sulphate from sulphate donor to galactose 6-hydroxy group etc. |
| XX | PT | |
| XX | PS | Claim 4; Page 15-16; 21pp; English. |
| XX | CC | This sequence produces the protein KSGal6ST (Keratan sulphate 6-sulphotransferase), which has a molecular weight of about 46700. The cDNA sequence of this protein was obtained by radiolabelling the cDNA of chick chondroitin 6-sulphotransferase (C6ST) and using this as a probe in a random oligonucleotide-primed labelling method. Human foetal brain cDNA was inserted into a phage Lambda gtl1 cloning vector whereby the clones containing the KSGal6ST were obtained by hybridization using the prepared probe. The positive clones were subcloned into a recombinant expression vector and used to transform COS-7 cells, from which cells expressing KSGal6ST can be selected. The phage cDNA inserts were isolated and subcloned into a Bluescript plasmid. Deletion clones were then prepared from which both strands were sequenced by the Sanger method. The KSGal6ST of the invention transfers the sulphate from a sulphate donor to galactose 6-OH groups in keratan sulphate, but does not transfer sulphate to chondroitin, chondroitin sulphate A or C, dermatan sulphate or CDSNs heparin |
| XX | CC | |
| XX | SQ | Sequence 411 AA; |
| | | Query Match 4.7%; Score 147; DB 2; Length 411; |
| | | Best Local Similarity 18.2%; Pred. No. 2.7e-05; |
| | | Matches 89; Conservative 58; Mismatches 149; Indels 194; Gaps 16 |
| Qy | 187 | WILLIVTALWFIEL-----LDVWSTC-----SQICAKWTTTEAGSKSLSS 229 |
| | | : : : : : : : : : : : : |
| Db | 5 | WKAVLLALASIAIQYTAIRFTTAKSFHTCFLGAEAGLAERLC-----EESPTFAYNLSR 59 |
| | | : : : : : : : : : : : : |
| Qy | 230 | EGHHMDLPDVVITSLPGSGAEILKOLFENSSDFLYIRVPTAVIDIPETELEIDSFVDACE 289 |
| | | : : : : : : : : : : : : |
| Db | 60 | KTH-----ILILATRRSGSSFGVQLFNQHLVDVFLFPEL----- 93 |
| | | : : : : : : : : : : : : |
| Qy | 290 | WKVSDIRSGHPRLLRGWILQSLVQDQTKLHLQNIHLHEPNRGKLAQVFANMKDKRKFRRR 349 |

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2005, 08:34:49 ; Search time 22.4643 Seconds
(without alignments)
1980.512 Million cell updates/sec

Title: US-10-697-828-9

Perfect score: 3147

Sequence: 1 MNRNGAMVDVDAHYKMF.....LIENICWTLMDRLGYPKFMD 596

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents AA.*
- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
 - 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
 - 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
 - 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
 - 5: /cgn2_6/ptodata/1/iaa/PCUS_COMB.pep.*
 - 6: /cgn2_6/ptodata/1/iaa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|----------------------|
| 1 | 568 | 18.0 | 126 | 4 | US-09-288-143-71 |
| 2 | 147 | 4.7 | 411 | 3 | US-09-015-188-2 |
| 3 | 134 | 4.3 | 483 | 3 | US-09-263-023-2 |
| 4 | 134 | 4.3 | 483 | 4 | US-09-471-867-2 |
| 5 | 133 | 4.2 | 484 | 3 | US-09-263-023-4 |
| 6 | 133 | 4.2 | 484 | 4 | US-09-471-867-4 |
| 7 | 133 | 4.2 | 531 | 4 | US-09-949-016-6471 |
| 8 | 133 | 4.2 | 608 | 4 | US-09-949-016-9449 |
| 9 | 123.5 | 3.9 | 395 | 4 | US-09-949-016-7011 |
| 10 | 120.5 | 3.8 | 479 | 2 | US-08-899-514-2 |
| 11 | 119.5 | 3.8 | 390 | 4 | US-09-949-016-6813 |
| 12 | 119.5 | 3.8 | 431 | 4 | US-09-949-016-8893 |
| 13 | 115.5 | 3.7 | 386 | 4 | US-09-786-240-11 |
| 14 | 108.5 | 3.4 | 386 | 3 | US-09-045-284A-2 |
| 15 | 108.5 | 3.4 | 386 | 3 | US-09-190-911-1 |
| 16 | 108 | 3.4 | 826 | 4 | US-09-252-991A-22143 |
| 17 | 103.5 | 3.3 | 324 | 4 | US-09-270-767-43224 |
| 18 | 103 | 3.3 | 582 | 3 | US-09-091-725-17 |
| 19 | 101.5 | 3.2 | 594 | 3 | US-08-987-151-2 |
| 20 | 101 | 3.2 | 650 | 4 | US-09-248-796A-16387 |
| 21 | 99.5 | 3.2 | 1060 | 4 | US-09-248-796A-15476 |
| 22 | 99.5 | 3.2 | 601 | 4 | US-09-107-433-3186 |
| 23 | 99 | 3.1 | 376 | 3 | US-09-150-133-7 |
| 24 | 99 | 3.1 | 376 | 3 | US-09-150-141-7 |
| 25 | 99 | 3.1 | 376 | 3 | US-09-374-493-7 |
| 26 | 99 | 3.1 | 376 | 3 | US-09-374-824-7 |
| 27 | 99 | 3.1 | 376 | 3 | US-09-374-492-7 |

| | | | | | | |
|----|------|-----|------|---|----------------------|-------------------|
| 28 | 99 | 3.1 | 376 | 4 | US-09-785-343-7 | Sequence 7, Appli |
| 29 | 99 | 3.1 | 376 | 4 | US-10-411-976-7 | Sequence 7, Appli |
| 30 | 98 | 3.1 | 591 | 4 | US-09-583-110-3756 | Sequence 3756, Ap |
| 31 | 97.5 | 3.1 | 1172 | 4 | US-09-328-352-6071 | Sequence 6071, Ap |
| 32 | 97 | 3.1 | 420 | 4 | US-09-949-016-6705 | Sequence 6705, Ap |
| 33 | 97 | 3.1 | 776 | 4 | US-09-949-016-7231 | Sequence 7231, Ap |
| 34 | 96.5 | 3.1 | 843 | 4 | US-09-248-796A-17210 | Sequence 17210, A |
| 35 | 96 | 3.1 | 458 | 2 | US-08-655-878-2 | Sequence 2, Appli |
| 36 | 94 | 3.0 | 232 | 4 | US-09-248-796A-16863 | Sequence 16863, A |
| 37 | 94 | 3.0 | 421 | 4 | US-09-949-016-8165 | Sequence 8165, Ap |
| 38 | 94 | 3.0 | 756 | 4 | US-09-919-497-66 | Sequence 66, Appl |
| 39 | 93 | 3.0 | 888 | 4 | US-09-538-092-787 | Sequence 787, App |
| 40 | 91.5 | 2.9 | 721 | 4 | US-08-851-435-2 | Sequence 2, Appli |
| 41 | 91.5 | 2.9 | 746 | 4 | US-08-851-435-6 | Sequence 6, Appli |
| 42 | 91.5 | 2.9 | 915 | 4 | US-09-583-110-3100 | Sequence 3100, Ap |
| 43 | 91 | 2.9 | 365 | 4 | US-09-818-512-4 | Sequence 4, Appli |
| 44 | 91 | 2.9 | 370 | 3 | US-09-150-133-1 | Sequence 1, Appli |
| 45 | 91 | 2.9 | 370 | 3 | US-09-150-141-1 | Sequence 1, Appli |

ALIGNMENTS

RESULT 1
US-09-288-143-71
; Sequence 71, Application US/09288143
; Patent No. 6433139
; GENERAL INFORMATION:
; APPLICANT: Brewer et al.
; TITLE OF INVENTION: 53 Human Secreted Proteins
; FILE REFERENCE: P2018P1
; CURRENT APPLICATION NUMBER: US/09/288,143
; CURRENT FILING DATE: 1999-04-08
; EARLIER APPLICATION NUMBER: PCT/US98/21142
; EARLIER FILING DATE: 1998-10-08
; EARLIER APPLICATION NUMBER: 60/061,463
; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: 60/061,529
; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: 60/071,498
; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: 60/061,527
; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: 60/061,536
; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: 60/061,532
; EARLIER FILING DATE: 1997-10-09
; NUMBER OF SEQ ID NOS: 219
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 71
; LENGTH: 126
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (126)
; OTHER INFORMATION: Xaa equals stop translation
US-09-288-143-71

Query Match 18.0%; Score 568; DB 4; Length 126;
Best Local Similarity 100.0%; Pred. No. 2e-53;
Matches 107; Conservative 0; Mismatches 0; Gaps 0;

| | | |
|----|-----|---|
| Qy | 185 | MRWLLIIVIALWFLFDVWSTCSQPCAKWTRTEAGSKLSSEGHMDLPDVTITSL 244 |
| Db | 1 | MRWLLIIVIALWFLFDVWSTCSQPCAKWTRTEAGSKLSSEGHMDLPDVTITSL 60 |
| Qy | 245 | PGSGAEILKQLFFNSSDFLIRVPTAYIDIPETELIDSFVDACEWK 291 |
| Db | 61 | PGSGAEILKQLFFNSSDFLIRVPTAYIDIPETELIDSFVDACEWK 107 |

RESULT 2

```

US-09-015-188-2
; FILE REFERENCE: US09015188C
; Sequence 2, Application US/09015188C
; Patent No. 6399358
; GENERAL INFORMATION:
; APPLICANT: Williams, Kevin J
; APPLICANT: Tabas, Ira
; TITLE OF INVENTION: A Human Gene Encoding Human Chondroitin
; TITLE OF INVENTION: 6-Sulfotransferase
; FILE REFERENCE: JEFF-0231
; CURRENT APPLICATION NUMBER: US/09/015,188C
; CURRENT FILING DATE: 1998-01-29
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 411
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-015-188-2

Query Match      4.7%; Score 147; DB 3; Length 411;
Best Local Similarity 18.2%; Pred. No. 9.6e-07;
Matches 89; Conservative 58; Mismatches 149; Indels 194; Gaps 16;

QY   187 WLLILVIALWFTEL-----LDVNSTC-----SQPICAKWTRTEAESKSLSS 229
       |||:|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db    5 WKAVLLALLASIAIQTAIRTFKTSFTFCGLAAGLAERLC-----EESPTFAYNLSR 59
       |||:|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

QY   230 EGHMDLPDVIITSLPGSGAEIKOLFNSSDLYIRVPTAYIDIPETELEIDSFYVDACE 289
       |||:|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db    60 KTH-----ILLIATRRGSSVFQGFNQHLDFVLFEPL-----93
       |||:|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

QY   290 WKVSDIRSGHFRLLRGWLQSLVODTKLHLQNHLHEPNRGKLAQYFAMNKDKRKPKRRE 349
       |||:|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db    94 -----YHVQNTLIPTFTQG-----KSPADRV 115
       |||:|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

QY   350 SUEPQRSMKGAFDRDAEVI-----RAURRHVVYPSPRPV-----LS 387
       |||:|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db   116 MLGASRDLLRSYDCDLVFLFNKYPPPVNHTDRIFRGASRVLCSPVCDPPGPADLV 175
       |||:|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

QY   388 LSSGSWTLKHFQFVLGR-----RLRWLYGTGR-----KEYNLDVTLQ-TTVCEDF 279
       |||:|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db   176 LEEGDCVRKCGLLNLTVAABACRSHSVAIKTVRVPEVNDLRALVEDPRNLKVQLVRD 235
       |||:|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

QY   417 PRAWIYSMLYNKPSLSYSLKNVPHEHLAKFKIEGGKGCNLSNGYAEPYELRKELSKSK 476
       |||:|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db   236 PRGILASRSETFDIY-----RLRWLYGTGR-----KEYNLDVTLQ-TTVCEDF 279
       |||:|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

QY   477 SNAYS--LLSHLWLANTAARINTDLPTSQYLKPEDIVHPQKTERIFAPLGIPL- 533
       |||:|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db   280 SNSVSTGLMRPPLXG-----KYMVLRYVEDLARNPMPKTEBIYGFLGIPLD 325
       |||:|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

QY   534 -----SPASLNQILPATSTNLFPYVEGEISPTNNVKQNLPRDEIKLIENI 581
       |||:|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db   326 SHVARMIQNTGDPDTLGHKHGYTVRN-----SAATAEKWRPRLSYDIVAFAQNA 375
       |||:|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

QY   582 CWTLMRLCY 591
       |||:|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db   376 CQQVLAQLGY 385
       |||:|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

RESULT 3
US-09-263-023-2
; Sequence 2, Application US/09263023
; Patent No. 6037159
; GENERAL INFORMATION:
; APPLICANT: Uchimura, Kenji
; APPLICANT: Muramatsu, Hideki
; APPLICANT: Kadomatsu, Kenji
; APPLICANT: Kannagi, Reiji
; APPLICANT: Habuchi, Osami
; APPLICANT: Muramatsu, Takashi
; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-O-SULFOTRANSFERASE AND
; TITLE OF INVENTION: DNA ENCODING THE SAME

US-09-015-188-2
; FILE REFERENCE: TOYAMA1.001AUS
; CURRENT APPLICATION NUMBER: US/09/263,023
; CURRENT FILING DATE: 1999-03-05
; EARLIER APPLICATION NUMBER: JP 10-54007
; EARLIER FILING DATE: 1998-03-05
; EARLIER APPLICATION NUMBER: JP 10-177844
; EARLIER FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 483
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-263-023-2

Query Match      4.3%; Score 134; DB 3; Length 483;
Best Local Similarity 26.2%; Pred. No. 3.3e-05;
Matches 53; Conservative 36; Mismatches 81; Indels 32; Gaps 9;

QY   408 MRALYIVRDPRAWIYSMLYNKPSLY--SLKNYPE-----HLAKLPKIEGGK-GKCNLNS 459
       |||:|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db   278 LKVTHLVRDPPRA-VASSRIRSRHGILRESLQVVRSDPPRAHRMPPFLEAAAGHKLGAKKEGM 336
       |||:|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

QY   460 GYAEPYEPLRKELSKSNVSVLLSHLWLANTAARINTDLPTSQYLKPEDIVHPQ 519
       |||:|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db   337 CGPADYHAL-----GAMEVICMSMATKTQTALQ-PEDWIQGHYLVVRYEDLVGDVP 386
       |||:|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

QY   530 KTTERRIFAFGLGPLSPA----SLNQILFATSTNLFPYVEGEISPTN----TNVMKQNL 571
       |||:|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db   387 KTLRRVDFVGLVSPNEHQFALNMVTSGSSGSKPFV-----VSARNATQAANAWTALT 441
       |||:|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

QY   572 RDEIKLIENICWTLMRLGYPK 593
       |||:|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db   442 FQIKQVEEFCYQPMAVLGYER 463
       |||:|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

RESULT 4
US-09-471-867-2
; Sequence 2, Application US/09471867
; Patent No. 6455289
; GENERAL INFORMATION:
; APPLICANT: Uchimura, Kenji
; APPLICANT: Muramatsu, Hideki
; APPLICANT: Kadomatsu, Kenji
; APPLICANT: Kannagi, Reiji
; APPLICANT: Habuchi, Osami
; APPLICANT: Muramatsu, Takashi
; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLVGLUCOSAMINE-6-O-SULFOTRANSFERASE AND
; TITLE OF INVENTION: DNA ENCODING THE SAME
; FILE REFERENCE: TOYAMA1.001AUS
; CURRENT APPLICATION NUMBER: US/09/471,867
; CURRENT FILING DATE: 1999-12-23
; EARLIER APPLICATION NUMBER: US 09/263,023
; EARLIER FILING DATE: 1999-03-05
; EARLIER APPLICATION NUMBER: JP 10-54007
; EARLIER FILING DATE: 1998-03-05
; EARLIER APPLICATION NUMBER: JP 10-177844
; EARLIER FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 483
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-471-867-2

Query Match      4.3%; Score 134; DB 4; Length 483;
Best Local Similarity 26.2%; Pred. No. 3.3e-05;
Matches 53; Conservative 36; Mismatches 81; Indels 32; Gaps 9;

QY   408 MRALYIVRDPRAWIYSMLYNKPSLY--SLKNYPE-----HLAKLPKIEGGK-GKCNLNS 459
       |||:|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db   278 LKVTHLVRDPPRA-VASSRIRSRHGILRESLQVVRSDPPRAHRMPPFLEAAAGHKLGAKKEGM 336
       |||:|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

```

```
; APPLICANT: Muramatsu, Takashi
; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-SULFOTRANSFERASE AND
; FILE REFERENCE: TOYAM41.001AUS
; CURRENT APPLICATION NUMBER: US/09/471,867
; CURRENT FILING DATE: 1999-12-23
; EARLIER APPLICATION NUMBER: US 09/263,023
; EARLIER FILING DATE: 1999-03-05
; EARLIER APPLICATION NUMBER: JP 10-54007
; EARLIER FILING DATE: 1998-03-05
; EARLIER APPLICATION NUMBER: JP 10-177844
; EARLIER FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Homo sapiens
; ORGANISM: Homo sapiens
US-09-471-867-4

Query Match 4.2%; Score 133; DB 3; Length 484;
Best Local Similarity 26.2%; Pred. No. 4.3e-05;
Matches 53; Conservative 36; Mismatches 81; Indels 32; Gaps 9;

Qy 408 MRALYIVRDPRAWIYMLYNSKPSLY--SLKNVPE-----HLAKLFKIEGK-GKCNLNS 459
Db 279 LKVIHLVRDPRA-VASSRIRSRHGLIRESLQVVRSDPRAHRMPPFLEAAGHKLGAKEGV 337
Qy 460 GYAFEYELPKELSKSNVSLSHLWNTAAALRINTDLPSTYQLVKFEFIVHFPQ 519
Db 338 GGPADYHAL-----GAMEVICNSMAKTLOTALQ-PPDWLQGHVLYVRYEDLVGDPV 387
Qy 520 KTERIFAFGLIPLSA-----SLNQILFATSTNLFYPYEGEISPTN-----TNVWKQNL 571
Db 388 KTLRRVDFVGLLVSPMEQFALNMTSGSGSSKPFV-----VSARNATQAANAWRTALT 442
Qy 572 RDEIKLIENICWTMLDRLGYPK 593
Db 443 FQIKQVEEFYQPMVILGYER 464

RESULT 7
US-09-949-016-6471
; Sequence 6471, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6471
; LENGTH: 531
; TYPE: PRT
; ORGANISM: Human
; ORGANISM: Human
US-09-949-016-6471

Query Match 4.2%; Score 133; DB 4; Length 531;
Best Local Similarity 26.2%; Pred. No. 5e-05;
Matches 53; Conservative 36; Mismatches 81; Indels 32; Gaps 9;

Qy 408 MRALYIVRDPRAWIYMLYNSKPSLY--SLKNVPE-----HLAKLFKIEGK-GKCNLNS 459
Db 326 LKVIHLVRDPRA-VASSRIRSRHGLIRESLQVVRSDPRAHRMPPFLEAAGHKLGAKEGV 384
```

```
; APPLICANT: Muramatsu, Takashi
; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-SULFOTRANSFERASE AND
; FILE REFERENCE: TOYAM41.001AUS
; CURRENT APPLICATION NUMBER: US/09/471,867
; CURRENT FILING DATE: 1999-12-23
; EARLIER APPLICATION NUMBER: US 09/263,023
; EARLIER FILING DATE: 1999-03-05
; EARLIER APPLICATION NUMBER: JP 10-54007
; EARLIER FILING DATE: 1998-03-05
; EARLIER APPLICATION NUMBER: JP 10-177844
; EARLIER FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Homo sapiens
; ORGANISM: Homo sapiens
US-09-471-867-4

Query Match 4.2%; Score 133; DB 4; Length 484;
Best Local Similarity 26.2%; Pred. No. 4.3e-05;
Matches 53; Conservative 36; Mismatches 81; Indels 32; Gaps 9;

Qy 408 MRALYIVRDPRAWIYMLYNSKPSLY--SLKNVPE-----HLAKLFKIEGK-GKCNLNS 459
Db 279 LKVIHLVRDPRA-VASSRIRSRHGLIRESLQVVRSDPRAHRMPPFLEAAGHKLGAKEGV 337
Qy 460 GYAFEYELPKELSKSNVSLSHLWNTAAALRINTDLPSTYQLVKFEFIVHFPQ 519
Db 338 GGPADYHAL-----GAMEVICNSMAKTLOTALQ-PPDWLQGHVLYVRYEDLVGDPV 387
Qy 520 KTERIFAFGLIPLSA-----SLNQILFATSTNLFYPYEGEISPTN-----TNVWKQNL 571
Db 388 KTLRRVDFVGLLVSPMEQFALNMTSGSGSSKPFV-----VSARNATQAANAWRTALT 442
Qy 572 RDEIKLIENICWTMLDRLGYPK 593
Db 443 FQIKQVEEFYQPMVILGYER 464

RESULT 7
US-09-949-016-6471
; Sequence 6471, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6471
; LENGTH: 531
; TYPE: PRT
; ORGANISM: Human
; ORGANISM: Human
US-09-949-016-6471

Query Match 4.2%; Score 133; DB 4; Length 531;
Best Local Similarity 26.2%; Pred. No. 5e-05;
Matches 53; Conservative 36; Mismatches 81; Indels 32; Gaps 9;

Qy 408 MRALYIVRDPRAWIYMLYNSKPSLY--SLKNVPE-----HLAKLFKIEGK-GKCNLNS 459
Db 326 LKVIHLVRDPRA-VASSRIRSRHGLIRESLQVVRSDPRAHRMPPFLEAAGHKLGAKEGV 384
```



```

; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: DANIEL E. ALTMAN
; REGISTRATION NUMBER: 34,115
; REFERENCE/DOCKET NUMBER: TOYAM21.001AUS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 714 760 0404
; TELEFAX: 714 760 9502
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 479
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-899-514-2

Query Match      3.8%; Score 120.5; DB 2; Length 479;
Best Local Similarity 16.7%; Pred. No. 0.00096;
Matches 71; Conservative 64; Mismatches 152; Indels 137; Gaps 12;

QY 217 RTEAGSKKSLSSGHHMDLPDVVITSLPGSGAEILKQLFFNSDDFLYIRVP----- 268
DB 118 RKEEPPRPVAGPRH-----VLLMATRTGSSFGVEFFNQOQNIPLYLPEPLMHIERTV 172
QY 269 -----TAYIDIPELEIDSFVDACEWKVSDIRSGH---FLLRGWLQSLVQ 312
DB 173 SFEPGGANAAGSALVYRDVLKQLFLCLDYV--LEHFTPLPEDHLTQPMFRGSRSLCE 230
QY 313 D-----TKLHLQNIHLHEPNRGKLAQYFAMNKDKKRRKRESLPQORSOMKGAFFRDA 366
DB 231 DPVCTPPVKVFEKYCHCKNRCGPNLTVLAAEACRRK-----EHWALKAVRIQOL 280
QY 367 EYRALRHLYVYPSAPVLSLSSGWTCLKLHFQEVGLGSMRALYIVRDPRAWIYMWLY 426
DB 281 EFLQPL-----AEDPRLDL-----RVQLVRDPRAVLASRMV 312
QY 427 NSKPSLYLKNVPEHLAKLFKEGGKGNLNSGYAFEPYELRKLKSKSKSNVLSLLSHL 486
DB 313 -----APAGKYTKWKWLDDGODGUREEVQ 339
QY 487 WLANTAALRINTDL-----LPTSQYLVKFEDIVHFPQKTTERRIFAFGLIPSLPASLN 539
DB 340 RLRCNCESIRLSAELGRQPAWLGRYVLYRYEDVARGPLQKAREMYFFAGIPLTPQVED 399
QY 540 QI-----LFTATNLFYLYPEGEISPTNTNWKQNLPRDEIKLIENICWTLMMD 587
DB 400 WIQNTQAAHDGSGIYSTQKN-----SSQFEKWFRESMPFKLAQVVQAPCGPAMR 449
QY 588 RLGY 591
DB 450 LFYG 453

RESULT 11
US-09-949-016-6813
; Sequence 6813, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0

Query Match      3.8%; Score 119.5; DB 4; Length 431;
Best Local Similarity 21.6%; Pred. No. 0.001;
Matches 86; Conservative 54; Mismatches 163; Indels 95; Gaps 16;

QY 223 SKKSLSEGHHMDLPDVVITSLPGSGAEILKQLFFNSDDFLYIRVP-----TA 270
DB 69 SRPGSPAGGEDRVHVLVLSWSRSGSFLGQFSDHPDVFYLMPEPAHWVTTLSQSSAA 128
QY 271 YIDIPETELEIDSFVDACEWKVSDI---RSGHFRLLRGWLQSLVQDTKLHLQNIHLHEP- 326

```

```

; SEQ ID NO 6813
; LENGTH: 390
; TYPE: PRT
; ORGANISM: Human
; US-09-949-016-6813

Query Match      3.8%; Score 119.5; DB 4; Length 390;
Best Local Similarity 21.6%; Pred. No. 0.00086;
Matches 86; Conservative 54; Mismatches 163; Indels 95; Gaps 16;

QY 223 SKKSLSEGHHMDLPDVVITSLPGSGAEILKQLFFNSDDFLYIRVP-----TA 270
DB 28 SRPGSPAGGEDRVHVLVLSWSRSGSFLGQFSDHPDVFYLMPEPAHWVTTLSQSSAA 87
QY 271 YIDIPETELEIDSFVDACEWKVSDI---RSGHFRLLRGWLQSLVQDTKLHLQNIHLHEP- 326
DB 88 TLHMAVRDLMRISFL--CMDVFDAYMPQSRNLSAFENWATSRA-----LCSPP 134
QY 327 -----NRGKLAQYFAMNKDKKRRKRESLPQORSOMKGAFFRDAEYIRALRHLYVYPS 381
DB 135 ACSAPFRGTISK-----QDVCKTLCTRQPFSLAREACRSYSHVVVLKEVRFPFNLOVLYPLL 189
QY 382 ARPVLSSSGSWTLKLHFFQEVGLGSMRALYIVRDPRAWIYMWLYSKNVPKH 441
DB 190 SDPALNL-----RIVHLVRDPRAVLRSR-EAAGPILARDNGI--- 225
QY 442 LAKLFKEGGKGNLNSGYAFEPY---LRKELSKSKSNVLSLLSHLWLAATAALRN 498
DB 226 -----VLGTNGK-----VEADPHLRILREVCR-----SHVRIAE-AATLK-P 261
QY 499 TDLPTSQYLVKFEDIVHFPQKTTERRIFAFGLIPSLPASLNQILFATSTNLFYLYPEG-E 557
DB 262 PPLRGYRLVRPDELAREPLAEIRALYAFGLTTLTPQLEAWIHNITHGSGIGKPIEAFH 321
QY 558 ISPTN-----TNWKQNLPRDEIKLIENICWTLMMDRLGY 591
DB 322 TSSRNARNVSNQAWRHALPFTKILRVQEVCAQALQLLGY 359

RESULT 12
US-09-949-016-8893
; Sequence 8893, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0

Query Match      3.8%; Score 119.5; DB 4; Length 431;
Best Local Similarity 21.6%; Pred. No. 0.001;
Matches 86; Conservative 54; Mismatches 163; Indels 95; Gaps 16;

QY 223 SKKSLSEGHHMDLPDVVITSLPGSGAEILKQLFFNSDDFLYIRVP-----TA 270
DB 69 SRPGSPAGGEDRVHVLVLSWSRSGSFLGQFSDHPDVFYLMPEPAHWVTTLSQSSAA 128
QY 271 YIDIPETELEIDSFVDACEWKVSDI---RSGHFRLLRGWLQSLVQDTKLHLQNIHLHEP- 326

```

[illegible]

```

Db      184  LQSLYPLKDPDS---LNLHI-----VHLVRDPA-----VFRSR----- 211
Qy      439  PEHLAKLFKIEGKGKCNLNSGYAF--EYPLRKELSKSKSNVSLLSHLWLANTAALR 496
Db      215  -----ERTKGLMIDSRIVMGQHEQKJKEKQDPYVMQVICQSQLEIYKTIQSL- 263
Qy      497  INTDILLPTSQYLKPEDIVHPQKTTTERRIFAFGLIPLSPASLNQILPAT-----STNLFY 551
Db      264  --PKALQERYLLVRVEDLARAPVAQTSRMTEFVGLGLEPLPLQTWVHNITRKGKMGDPAHF 321
Qy      552  LPEVEGEISPTN-----TNVWKONLPRDEIKLIENICWTLMRLGY 591
Db      322  -----TNARDALNVSQAWRWSLPYEKVRSRLQKACGDMNLLGY 359

RESULT 14
US-09-045-284A-2
; Sequence 2, Application US/09045284A
; Patent No. 6285192
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107051
; CURRENT APPLICATION NUMBER: US/09/045,284A
; CURRENT FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-045-284A-2

Query Match          3.4%; Score 108.5; DB 3; Length 386;
Best Local Similarity 18.9%; Pred. No. 0.013;
Matches 78; Conservative 55; Mismatches 151; Indels 129; Gaps 15

Qy      223  SKKSLSSGHHMDLPDVVITSLPGSGAEILKQLFFNSSDFLYIRVPTAVIDIPETELEID 282
Db      32  SLSLWKAQPERK---HVLVLSSWRSGSFFVGQULFGQHPDVFYLMERPAHV----- 78
Qy      283  SPVDACEKWKVSDIRSGHFRLLRGWLQSLAVQDQTKLHLQN IHLHPEPNRGKLAQYFAMNKDKK 342
Db      79  -----WMT-----FKQSTAMW-----LHMAVRDLIRAVFLCDMSVP 109
Qy      343  RKFKRRRESLPGRSQMGAPDRDAEYTRAL-----RRHLVYPPSARPV 386
Db      110  DAY--MEPGPRQSSL-----FQWENSRALCSAPACDII PQDEIIIPRAHCRLLCSQOPFE 162
Qy      387  SLSSGSGWTLLKHFQEOVLGASMRALY-----IVRDPRAWIYSMLYNSKPSL 432
Db      163  VVEKACRSYSHVLUKEVRFFNLQSLYPLLPDKDPSLNHLHIVHLVRDPRA-----VFRSR----- 214
Qy      433  YSLKNVPEHLAKLFKIEGKGKCNLNSGYAF--EYEPRLKELSKSKSNVSLLSHLWLAN 490
Db      215  -----ERTKGLMIDSRIVMGQHEQKJKEKQDPYVMQVICQSQLEIYK 258
Qy      491  TAAALRINTDLLPTSQYLKPEDIVHPQKTTTERRIFAFGLIPLSPASLNQILPAT----- 545
Db      259  TTIQSL--PKALQERYLLVRVEDLARAPVAQTSRMTEFVGLGLEPLPLQTWVHNITRKGKMG 315
Qy      546  STNLFYLPVEGEISPTN-----TNVWKONLPRDEIKLIENICWTLMRLGY 591
Db      316  GDPAHF-----TNARDALNVSQAWRWSLPYEKVRSRLQKACGDMNLLGY 359

RESULT 15
US-09-190-911-1
; Sequence 1, Application US/091090911
; Patent No. 6365365
; GENERAL INFORMATION:

```

; APPLICANT: Bistrup, Annette
 ; APPLICANT: Rosen, Steven D.
 ; APPLICANT: Tangemann, Kirsten
 ; APPLICANT: Hemmerich, Stefan
 ; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
 ; FILE REFERENCE: 6510-107CIP
 ; CURRENT APPLICATION NUMBER: US/09/190,911
 ; CURRENT FILING DATE: 1998-11-12
 ; EARLIER APPLICATION NUMBER: 09/045,284
 ; EARLIER FILING DATE: 1998-03-20
 ; NUMBER OF SEQ ID NOS: 8
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 1
 ; LENGTH: 386
 ; TYPE: PRT
 ; ORGANISM: H. sapiens
 US-09-190-911-1

Query Match 3.4%; Score 108.5; DB 3; Length 386;
 Best Local Similarity 18.9%; Pred. No. 0.013;
 Matches 78; Conservative 55; Mismatches 151; Indels 129; Gaps 15;

| | | | |
|----|-----|---|-----|
| Qy | 223 | SKSLSEGHNDLPDVVITSLPGSGAILKQLFNSDDFLYIRVPTAYIDIPETELEID | 282 |
| Db | 32 | SSLGSKAQPERM---HVLVLSWRSGSFGQLFGQHPDVFYLMPEAHV----- | 78 |
| Qy | 283 | SFVDACEKVS DIRSGHFRLLRGLQSLVQDTKLHLQNIHLHFNRGKLAQYFAMNKDKK | 342 |
| Db | 79 | -----WMT-----FKQSTAWM-----LHMAVRDLIRAVFLCDMSVF | 109 |
| Qy | 343 | RKFRRESLPQRSQMGAFDRDAEYIRAL-----RRHLVYYPSPARPVL | 386 |
| Db | 110 | DAY--MEGPRRQSSL-----FQWENSRALCSAPACDIIPODEIIPRAHCRLLCSQQPFE | 162 |
| Qy | 387 | SLSSGSWTLKLHFTQEVVLGASMRALY-----IVRDPRAWIYSMLYNSKPSL | 432 |
| Db | 163 | VVERACRSYSHVWLKEVRFFNLQSLYPLLKDPSLNLHIVHLVRDPA-----VFRSR--- | 214 |
| Qy | 433 | YSLKNVPEHLAKLFIEGKGKCNLNGYAF--EYEPLRKELSKSKSNVSLLSHLMLAN | 490 |
| Db | 215 | -----ERTKGLMIDSRIVMGQHEQKLKEDQPYVYVQVVCQSQLEIYK | 258 |
| Qy | 491 | TAAALRINTDLLPTSQYLVKPEDIVHFPQKTERIFAFGLPGLSPASLNQILFAT----- | 545 |
| Db | 259 | TIQSL---PKALQERYLLVRYEDLARAPVAQTSRMVEFVGLFLEFLHLQTVHNI TRCKGM | 315 |
| Qy | 546 | STNLFYLPYEGEISFTN-----TNWVKQLPRDEIKLIENICWTLMDRIGY | 591 |
| Db | 316 | GDHAFH-----TNARDALNVSAWRWSLPYEKVSRLQKACGDMNLLGY | 359 |

Search completed: June 23, 2005, 08:52:13
 Job time : 24.4643 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2005, 08:49:14 ; Search time 75.3449 Seconds
(without alignment)

3041.886 Million cell updates/sec

Title: US-10-697-828-9

Perfect score: 3147

Sequence: 1 MNRVNGAMDVWDVDAHYKMFV.....LIENICWTLMDRLGYPKMD 596

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 171557 seqs, 384547976 residues

Total number of hits satisfying chosen parameters: 1717557

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/1/pubpaa/US11A_PUBCOMB.pep.*
- 20: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
- 21: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 22: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB ID | Description |
|------------|-------|---------------|--------|-------|--------------------|
| 1 | 3147 | 100.0 | 596 | 16 | US-10-697-828-9 |
| 2 | 3147 | 100.0 | 1222 | 16 | US-10-697-828-15 |
| 3 | 3142 | 99.8 | 1212 | 17 | US-10-479-472A-2 |
| 4 | 3142 | 99.8 | 1222 | 16 | US-10-475-446-4 |
| 5 | 2835 | 90.1 | 1207 | 17 | US-10-697-828-17 |
| 6 | 2835 | 90.1 | 1207 | 16 | US-10-479-472A-4 |
| 7 | 568 | 18.0 | 125 | 11 | US-09-984-429-71 |
| 8 | 568 | 18.0 | 126 | 14 | US-10-150-111-71 |
| 9 | 147 | 4.7 | 411 | 14 | US-10-021-660-128 |
| 10 | 147 | 4.7 | 411 | 15 | US-10-211-462-97 |
| 11 | 147 | 4.7 | 411 | 16 | US-10-408-765A-395 |

| | | | | | | |
|----|-------|-----|------|----|----------------------|--------------------|
| 12 | 147 | 4.7 | 411 | 16 | US-10-723-860-1544 | Sequence 1544, Ap |
| 13 | 134 | 4.3 | 483 | 14 | US-10-212-933-2 | Sequence 2, Appli |
| 14 | 133 | 4.2 | 484 | 14 | US-10-212-933-4 | Sequence 4, Appli |
| 15 | 133 | 4.2 | 530 | 16 | US-10-723-860-1409 | Sequence 1409, Ap |
| 16 | 133 | 4.2 | 531 | 9 | US-09-833-790-255 | Sequence 255, App |
| 17 | 133 | 4.2 | 531 | 16 | US-10-755-889-284 | Sequence 284, App |
| 18 | 123.5 | 3.9 | 395 | 9 | US-09-927-602-2 | Sequence 2, Appli |
| 19 | 123.5 | 3.9 | 395 | 15 | US-10-258-080-5 | Sequence 5, Appli |
| 20 | 123.5 | 3.9 | 395 | 16 | US-10-648-593-159 | Sequence 159, App |
| 21 | 123.5 | 3.9 | 395 | 16 | US-10-697-828-13 | Sequence 13, Appl |
| 22 | 122.5 | 3.9 | 479 | 13 | US-10-087-192-126 | Sequence 126, App |
| 23 | 119.5 | 3.8 | 390 | 9 | US-09-927-602-4 | Sequence 4, Appli |
| 24 | 119.5 | 3.8 | 390 | 16 | US-10-697-828-8 | Sequence 8, Appli |
| 25 | 119 | 3.8 | 481 | 13 | US-10-087-192-123 | Sequence 123, App |
| 26 | 118.5 | 3.8 | 418 | 9 | US-09-927-602-5 | Sequence 5, Appli |
| 27 | 118 | 3.7 | 395 | 16 | US-10-697-828-7 | Sequence 7, Appli |
| 28 | 118 | 3.7 | 395 | 16 | US-10-841-707-8 | Sequence 8, Appli |
| 29 | 115.5 | 3.7 | 386 | 14 | US-10-427-631-11 | Sequence 11, Appl |
| 30 | 113 | 3.6 | 395 | 9 | US-09-927-602-3 | Sequence 3, Appli |
| 31 | 109.5 | 3.5 | 1441 | 15 | US-10-346-863-6 | Sequence 6, Appli |
| 32 | 109.5 | 3.5 | 1441 | 16 | US-10-408-765A-824 | Sequence 824, App |
| 33 | 108.5 | 3.4 | 380 | 16 | US-10-841-707-6 | Sequence 6, Appli |
| 34 | 108.5 | 3.4 | 386 | 9 | US-09-816-825-2 | Sequence 2, Appli |
| 35 | 108.5 | 3.4 | 386 | 13 | US-10-007-262-1 | Sequence 1, Appli |
| 36 | 103 | 3.3 | 582 | 15 | US-10-001-192A-17 | Sequence 17, Appl |
| 37 | 102 | 3.2 | 1264 | 14 | US-10-032-585-7118 | Sequence 7118, Ap |
| 38 | 101.5 | 3.2 | 591 | 15 | US-10-282-122A-73955 | Sequence 73955, A |
| 39 | 100.5 | 3.2 | 703 | 16 | US-10-437-963-104263 | Sequence 104263, A |
| 40 | 100 | 3.2 | 1663 | 17 | US-10-732-923-13998 | Sequence 13998, A |
| 41 | 100 | 3.2 | 2322 | 17 | US-10-732-923-13997 | Sequence 13997, A |
| 42 | 99.5 | 3.2 | 590 | 17 | US-10-732-923-1842 | Sequence 1842, Ap |
| 43 | 99 | 3.1 | 376 | 14 | US-10-411-976-7 | Sequence 7, Appli |
| 44 | 99 | 3.1 | 519 | 15 | US-10-425-114-45870 | Sequence 45870, A |
| 45 | 99 | 3.1 | 2360 | 17 | US-10-732-923-13960 | Sequence 13960, A |

ALIGNMENTS

RESULT 1
US-10-697-828-9
; Sequence 9, Application US/10697828
; Publication No. US20040185546A1
; GENERAL INFORMATION:
; APPLICANT: Rosen, Steven
; APPLICANT: Lee, Jin Kyu
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: Novel Glycosyl Sulfotransferases GST-4alpha, GST-4beta, & GST-6
; FILE REFERENCE: UCAL-138DIV
; CURRENT APPLICATION NUMBER: US/10/697,828
; PRIOR FILING DATE: 2003-10-29
; PRIOR APPLICATION NUMBER: 09/593,828
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: 60/144,694
; PRIOR FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 596
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-697-828-9

Query Match 100.0%; Score 3147; DB 16; Length 596;
Best Local Similarity 100.0%; Pred. No. 2.5e-295;
Matches 596; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MNRVNGAMDVWDVDAHYKMFVFDHGNSPMASIQAEQAAEFKCKWTQFVNVTFQMESIT 60
|||||
Db 1 MNRVNGAMDVWDVDAHYKMFVFDHGNSPMASIQAEQAAEFKCKWTQFVNVTFQMESIT 60
|||||
Qy 61 RIAYVFGPYINVSSCRFIDSSNFGQLISLVNNTTEHVSVITVDYHNLKTRFNYLGFQGF 120
|||||

```
Db 61 RIAYVFGPYINSSCRFIDSSNPGLOISLVNNTTEHWSIVTDYHNLKTRFNVLGFGF 120
Qy 121 ASVADQOITRFGLGTOAIVKPVHRDRIIPFGFKFNIAVGLILCISLVILTFQWRPFLS 180
Db 121 ASVADQOITRFGLGTOAIVKPVHRDRIIPFGFKFNIAVGLILCISLVILTFQWRPFLS 180
Qy 181 FRKLMRWILILVIALWFIELLDVWSTCSQPICAKWTTRTEAGSKSLSSSEGHMDLPDVV 240
Db 181 FRKLMRWILILVIALWFIELLDVWSTCSQPICAKWTTRTEAGSKSLSSSEGHMDLPDVV 240
Qy 241 ITSLPGSGABILKQLFFNSDFLYIRVPTAYIDIPETELEIDSFVDAACEWKVSDIRSGHF 300
Db 241 ITSLPGSGABILKQLFFNSDFLYIRVPTAYIDIPETELEIDSFVDAACEWKVSDIRSGHF 300
Qy 301 RLLRGWLQSLVQDTKLHQLNIHLHEPNRGKLAQYFAMNKDKKFKRRESLPRQSQMKG 360
Db 301 RLLRGWLQSLVQDTKLHQLNIHLHEPNRGKLAQYFAMNKDKKFKRRESLPRQSQMKG 360
Qy 361 AFDRDAEYIRALRRHLVYPSARPVLSLSSGSWTLKLHFFQEVLGASMRALYIVRDPRAW 420
Db 361 AFDRDAEYIRALRRHLVYPSARPVLSLSSGSWTLKLHFFQEVLGASMRALYIVRDPRAW 420
Qy 421 IYSLMYSKPSLSLXKNVPEHLAKLFKIEGGKGCNLSGYAFEYEPRLKELSKSKSNAV 480
Db 421 IYSLMYSKPSLSLXKNVPEHLAKLFKIEGGKGCNLSGYAFEYEPRLKELSKSKSNAV 480
Qy 481 SLLSHLWLANTAALRINTDLLPTSQYLVKFPEDIVHPPOKTTTERIFAFGLGIPLSASLNQ 540
Db 481 SLLSHLWLANTAALRINTDLLPTSQYLVKFPEDIVHPPOKTTTERIFAFGLGIPLSASLNQ 540
Qy 541 ILPATSTNLFLPYEGEISPTNTNVWKQNLPRDEIKLIENICWTLMDRLGYPKFM 596
Db 541 ILPATSTNLFLPYEGEISPTNTNVWKQNLPRDEIKLIENICWTLMDRLGYPKFM 596

RESULT 2
; Sequence 15, Application US/10697828
; Publication No. US20040185546A1
; GENERAL INFORMATION:
; APPLICANT: Rosen, Steven
; APPLICANT: Lee, Jin Kyu
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: Novel Glycosyl Sulfotransferases GST-4alpha, GST-4beta, & GST-6
; FILE REFERENCE: UCAL-138Div
; CURRENT APPLICATION NUMBER: US/10/697,828
; PRIOR FILING DATE: 2003-10-29
; PRIOR APPLICATION NUMBER: 09/593,828
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: 60/144,694
; PRIOR FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 1222
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-697-828-15

Query Match 100.0%; Score 3147; DB 16; Length 1222;
Best Local Similarity 100.0%; Pred. No. 7, 6e-295;
Matches 596; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MNRINGAMMDVWDHYKMFWDHGHGNSPMASIQEAEQAAEFKKRWTQFVNVVTFQMESTIT 60
Db 627 MNRINGAMMDVWDHYKMFWDHGHGNSPMASIQEAEQAAEFKKRWTQFVNVVTFQMESTIT 686
Qy 61 RIAYVFGPYINSSCRFIDSSNPGLOISLVNNTTEHWSIVTDYHNLKTRFNVLGFGF 120
Db 687 RIAYVFGPYINSSCRFIDSSNPGLOISLVNNTTEHWSIVTDYHNLKTRFNVLGFGF 746
Qy 121 ASVADQOITRFGLGTOAIVKPVHRDRIIPFGFKFNIAVGLILCISLVILTFQWRPFLS 180
```

```
Db 747 ASVADQOITRFGLGTOAIVKPVHRDRIIPFGFKFNIAVGLILCISLVILTFQWRPFLS 806
Qy 181 FRKLMRWILILVIALWFIELLDVWSTCSQPICAKWTTRTEAGSKSLSSSEGHMDLPDVV 240
Db 807 FRKLMRWILILVIALWFIELLDVWSTCSQPICAKWTTRTEAGSKSLSSSEGHMDLPDVV 866
Qy 241 ITSLPGSGABILKQLFFNSDFLYIRVPTAYIDIPETELEIDSFVDAACEWKVSDIRSGHF 300
Db 867 ITSLPGSGABILKQLFFNSDFLYIRVPTAYIDIPETELEIDSFVDAACEWKVSDIRSGHF 926
Qy 301 RLLRGWLQSLVQDTKLHQLNIHLHEPNRGKLAQYFAMNKDKKFKRRESLPRQSQMKG 360
Db 927 RLLRGWLQSLVQDTKLHQLNIHLHEPNRGKLAQYFAMNKDKKFKRRESLPRQSQMKG 986
Qy 361 AFDRDAEYIRALRRHLVYPSARPVLSLSSGSWTLKLHFFQEVLGASMRALYIVRDPRAW 420
Db 987 AFDRDAEYIRALRRHLVYPSARPVLSLSSGSWTLKLHFFQEVLGASMRALYIVRDPRAW 1046
Qy 421 IYSLMYSKPSLSLXKNVPEHLAKLFKIEGGKGCNLSGYAFEYEPRLKELSKSKSNAV 480
Db 1047 IYSLMYSKPSLSLXKNVPEHLAKLFKIEGGKGCNLSGYAFEYEPRLKELSKSKSNAV 1106
Qy 481 SLLSHLWLANTAALRINTDLLPTSQYLVKFPEDIVHPPOKTTTERIFAFGLGIPLSASLNQ 540
Db 1107 SLLSHLWLANTAALRINTDLLPTSQYLVKFPEDIVHPPOKTTTERIFAFGLGIPLSASLNQ 1166
Qy 541 ILPATSTNLFLPYEGEISPTNTNVWKQNLPRDEIKLIENICWTLMDRLGYPKFM 596
Db 1167 ILPATSTNLFLPYEGEISPTNTNVWKQNLPRDEIKLIENICWTLMDRLGYPKFM 1222

RESULT 3
US-10-479-472A-2
; Sequence 2, Application US/10479472A
; Publication No. US20050118581A1
; GENERAL INFORMATION:
; APPLICANT: DEL-FAVERO, JURGEN PETER LOE
; APPLICANT: VAN BROECKHOVEN, CHRISTINE
; TITLE OF INVENTION: NOVEL BRAIN EXPRESSED GENE AND PROTEIN ASSOCIATED WITH
; TITLE OF INVENTION: BIPOLAR DISORDER
; FILE REFERENCE: JAB-1711
; CURRENT APPLICATION NUMBER: US/10/479,472A
; PRIOR FILING DATE: 2003-12-01
; PRIOR APPLICATION NUMBER: PCT/EP02/06316
; PRIOR FILING DATE: 2002-06-06
; PRIOR APPLICATION NUMBER: EP 01202214.1
; PRIOR FILING DATE: 2001-06-11
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 2
; LENGTH: 1212
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-479-472A-2

Query Match 99.8%; Score 3142; DB 17; Length 1212;
Best Local Similarity 99.8%; Pred. No. 2.3e-294;
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MNRINGAMMDVWDHYKMFWDHGHGNSPMASIQEAEQAAEFKKRWTQFVNVVTFQMESTIT 60
Db 617 MNRINGAMMDVWDHYKMFWDHGHGNSPMASIQEAEQAAEFKKRWTQFVNVVTFQMESTIT 676
Qy 61 RIAYVFGPYINSSCRFIDSSNPGLOISLVNNTTEHWSIVTDYHNLKTRFNVLGFGF 120
Db 677 RIAYVFGPYINSSCRFIDSSNPGLOISLVNNTTEHWSIVTDYHNLKTRFNVLGFGF 736
Qy 121 ASVADQOITRFGLGTOAIVKPVHRDRIIPFGFKFNIAVGLILCISLVILTFQWRPFLS 180
Db 737 ASVADQOITRFGLGTOAIVKPVHRDRIIPFGFKFNIAVGLILCISLVILTFQWRPFLS 796
Qy 181 FRKLMRWILILVIALWFIELLDVWSTCSQPICAKWTTRTEAGSKSLSSSEGHMDLPDVV 240
```

797 FRKLMRWILILVIALWFIELDDVWSTCSQPTCAKWTREAGSKSLSSSEGHMDLPDV 856
241 ITSLPGSGAEILKOLFNSDFLYIRVPTAYIDIPETELEIDSFVDCACWKVSDIRSGHF 300
857 ITSLPGSGAEILKOLFNSDFLYIRVPTAYIDIPETELEIDSFVDCACWKVSDIRSGHF 916
301 RLLRGWLQSLVQDTKLHLQNLHLEPNRGKLAQYFAMNKKRKFRRRESLPEQRSQMGK 360
917 RLLRGWLQSLVQDTKLHLQNLHLEPNRGKLAQYFAMNKKRKFRRRESLPEQRSQMGK 976
361 AFDRDAEYIRALRHLYVYPSARVLSLSSGSGWTCLKLHFFQEVLGASMRALYIYVRDPRAW 420
977 AFDRDAEYIRALRHLYVYPSARVLSLSSGSGWTCLKLHFFQEVLGASMRALYIYVRDPRAW 1036
421 IYSLMYSKPSLSYSLKNVPEHLAKLFKEGGKGCNLSNGVAFYEPLRKELSKSKSNV 480
1037 IYSLMYSKPSLSYSLKNVPEHLAKLFKEGGKGCNLSNGVAFYEPLRKELSKSKSNV 1096
481 SLLSHLWLANTAALRINTDLLPTSQYLVKPEDIVHFPQKTTERRIFAFGLGIPLSPASLNQ 540
1097 SLLSHLWLANTAALRINTDLLPTSQYLVKPEDIVHFPQKTTERRIFAFGLGIPLSPASLNQ 1156
541 ILFATSTNLFYLPYEGEISPTNTNWKNLPRDEIKLIENICWTLMDBRLGYPKFMD 596
1157 ILFATSTNLFYLPYEGEISPTNTNWKNLPRDEIKLIENICWTLMDBRLGYPKFMD 1212

RESULT 4

US-10-475-446-4
; Sequence 4, Application US/10475446
; Publication No. US20040198651A1
; GENERAL INFORMATION:
; APPLICANT: KLAMMER, Aaron A.; HAFALIA, April J.A.
; APPLICANT: DUGAN, Brendan M.; WARREN, Bridget A.
; APPLICANT: EMERLING, Brooke M.; TRIBOULEY, Catherine M.
; APPLICANT: ARVIZU, Chandra S.; HONCHELL, Cynthia D.
; APPLICANT: NGUYEN, Dannel B.; KALLICK, Deborah A.
; APPLICANT: YUE, Henry; AU-YOUNG, Janice K.
; APPLICANT: RAKUMAR, Jayalaxmi; LI, Joana X.
; APPLICANT: THANGAVARU, Kavitha; GIETZEN, Kimberly J.
; APPLICANT: DING, Li; BAUGHN, Mariah R.
; APPLICANT: YAO, Monique G.; CHAWLA, Narinder K.
; APPLICANT: MASON, Patricia M.; LAL, Preeti G.
; APPLICANT: GAUL, Richard C.; REDDY, Roopa M.
; APPLICANT: BECHA, Shanya D.; KAREHT, Stephanie K.
; APPLICANT: RICHARDSON, Thomas W.; TRAN, Uyen K.
; APPLICANT: ELLIOTT, Vicki S.; TANG, Y. Tom
; APPLICANT: AZIMZAI, Yalda; LU, Yan
; APPLICANT: XU, Yuming
; TITLE OF INVENTION: SECRETED PROTEINS
; FILE REFERENCE: PF-0949 USN
; CURRENT APPLICATION NUMBER: US/10/475,446
; PRIOR FILING DATE: 2003-10-20
; PRIOR APPLICATION NUMBER: PCT/US02/12464
; PRIOR FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US 60/285,207
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 60/287,114
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/288,640
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: US 60/290,516
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/292,184
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 60/343,553
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: US 60/357,002
; PRIOR FILING DATE: 2002-02-13
; PRIOR APPLICATION NUMBER: US 60/358,279
; PRIOR FILING DATE: 2002-02-20
; PRIOR APPLICATION NUMBER: US 60/366,041
; PRIOR FILING DATE: 2002-03-19

; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PERL Program
; SEQ ID NO 4
; LENGTH: 1222
; TYPE: PRM
; ORGANISM: Homo sapiens
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 3441255CD1
US-10-475-446-4

Query Match 99.8%; Score 3142; DB 16; Length 1222;
Best Local Similarity 99.8%; Pred. No. 2.3e-294;
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MNRVNGAMVDVDAHYKMFDFDHGNSPMASIOBAEQAAEFKKRWTFQVNVTFQWESFIT 60
Db 627 MNRVNGAMVDVDAHYKMFDFDHGNSPMASIOBAEQAAEFKKRWTFQVNVTFQWESFIT 686
Qy 61 RIAYVFYGPYINVSSCRFIDSSNPGLQISLVNNTNTEHVSVITVDYHNLKTRFNYLGGGF 120
Db 687 RIAYVFYGPYINVSSCRFIDSSNPGLQISLVNNTNTEHVSVITVDYHNLKTRFNYLGGGF 746
Qy 121 ASVADQSQITRFGLTQAIKVPVRHRIIPFGPKFNIAGVLICISLVILTQWRFYLS 180
Db 747 ASVADQSQITRFGLTQAIKVPVRHRIIPFGPKFNIAGVLICISLVILTQWRFYLS 806
Qy 181 FRKLMRWILILVIALWFIELDDVWSTCSQPTCAKWTREAGSKSLSSSEGHMDLPDV 240
Db 807 FRKLMRWILILVIALWFIELDDVWSTCSQPTCAKWTREAGSKSLSSSEGHMDLPDV 866
Qy 241 ITSLPGSGAEILKOLFNSDFLYIRVPTAYIDIPETELEIDSFVDCACWKVSDIRSGHF 300
Db 867 ITSLPGSGAEILKOLFNSDFLYIRVPTAYIDIPETELEIDSFVDCACWKVSDIRSGHF 926
Qy 301 RLLRGWLQSLVQDTKLHLQNLHLEPNRGKLAQYFAMNKKRKFRRRESLPEQRSQMGK 360
Db 927 RLLRGWLQSLVQDTKLHLQNLHLEPNRGKLAQYFAMNKKRKFRRRESLPEQRSQMGK 986
Qy 361 AFDRDAEYIRALRHLYVYPSARVLSLSSGSGWTCLKLHFFQEVLGASMRALYIYVRDPRAW 420
Db 987 AFDRDAEYIRALRHLYVYPSARVLSLSSGSGWTCLKLHFFQEVLGASMRALYIYVRDPRAW 1046
Qy 421 IYSLMYSKPSLSYSLKNVPEHLAKLFKEGGKGCNLSNGVAFYEPLRKELSKSKSNV 480
Db 1047 IYSLMYSKPSLSYSLKNVPEHLAKLFKEGGKGCNLSNGVAFYEPLRKELSKSKSNV 1106
Qy 481 SLLSHLWLANTAALRINTDLLPTSQYLVKPEDIVHFPQKTTERRIFAFGLGIPLSPASLNQ 540
Db 1107 SLLSHLWLANTAALRINTDLLPTSQYLVKPEDIVHFPQKTTERRIFAFGLGIPLSPASLNQ 1166
Qy 541 ILFATSTNLFYLPYEGEISPTNTNWKNLPRDEIKLIENICWTLMDBRLGYPKFMD 596
Db 1167 ILFATSTNLFYLPYEGEISPTNTNWKNLPRDEIKLIENICWTLMDBRLGYPKFMD 1222

RESULT 5

US-10-697-828-17
; Sequence 17, Application US/10697828
; Publication No. US20040185546A1
; GENERAL INFORMATION:
; APPLICANT: Rosen, Steven
; APPLICANT: Lee, Jin Kyu
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: Novel Glycosyl Sulfotransferases GST-4alpha, GST-4beta, & GST-6
; FILE REFERENCE: UCAL-138DIV
; CURRENT APPLICATION NUMBER: US/10/697,828
; CURRENT FILING DATE: 2003-10-29
; PRIOR APPLICATION NUMBER: 09/593,828
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: 60/144,694
; PRIOR FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 23

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 1207
; TYPE: PRT
; ORGANISM: mouse
US-10-697-828-17

Query Match          90.1%; Score 2835; DB 16; Length 1207;
Best Local Similarity 88.6%; Pred. No. 1.3e-264;
Matches 528; Conservative 36; Mismatches 32; Indels 0; Gaps 0;

QY 1 MNRNGAMDMVDADHYKMFDFDHGNSPMASIOEAEQAABFKKRWTFQVNVVTFQMESTIT 60
DB 612 MNRNGAMDMVDADHYKMFDFDHGNSPMVANOIEAEQAABFKKRWTFQVNVVTFHMESTIT 671

QY 61 RIAYVFGPYNVSSCRFDIDSSNPGIQISLVNNTTEHVSVIVTDYHNLKTRFNYLGGGFF 120
DB 672 RIAYVFGPYNVSSCRFDIDSSSGLQISLVNSTEHSVSVVTDYQNLKSRFSYLGFGGF 731

QY 121 ASVADOGITRFGIGTQAIKVPVRHDIRIIPFGFKFNIAVGLICISLVILTFQWRPYLS 180
DB 732 ASVANOQITRFGIGTQEIIVNVRHDKVNPFPFGFKFNIAVGFILCISLVILTFQWRPYLS 791

QY 181 FRKLMRWILVIALWFIELLDVWSTCSQPICAKWTTEAGSKSLSSSEGHMDLPDVV 240
DB 792 FRKLMRCVLIVIALWFIELLDVWSTCTQPICAKWTTEAKANEKVMISSEGHVVDLPNVI 851

QY 241 ITSLPGSGAEILKQLFNSDDFLYIRVPTAYIDIPETELEDISFVDACEWKVSDIRSGHF 300
DB 852 ITSLPGSGAEILKQLFNSDDFLYIRIPTAYMDIPETELEDISFVDACEWKVSDIRSGHF 911

QY 301 RLRLGWLQSLVQDTKLHLQNIHLHFNPKLAQYFAMNKDKKFKRRESLPSQSRQMGK 360
DB 912 HLLRGWLQSLVQDTKLHLQNIHLHETSRKLAQYFTTNKDKKFKRRESLQDQSRKIG 971

QY 361 AFDRDAEYIRALRRHLVYPSARPVLSSSGSWTLKLHFFQEVLGASMRALYIVRDPRAW 420
DB 972 PFDRDAEYIRALRRHLVYPSARPVLSSSGSWTLKLHFFQEVLGTSMRALYIVRDPRAW 1031

QY 421 IYSMLYNSKPSLSYSLKNVPEHLAKLFKIEGGKGCNLSGYAFEPYELPKELSKSNAV 480
DB 1032 IYSVLGSKPSLSYSLKNVPEHLAKLFKIEGSKCNSNSGYAFEYESLKELETSQSNAI 1091

QY 481 SLLSHLWLANTAALRINTDLPSTYQLVKFEDIIVHPQKTTERRIFAFGLGIPLSPASLNQ 540
DB 1092 SLLSHLWVANTAALRINTDLPNTYHLVKFEDIVHPQKTTERRIFAFGLGIPLSPASLNQ 1151

QY 541 ILFATSTNLFYLPYEGEISPTNTNVKQNLPRDEIKLIENICWTLMDRLGYPKFMD 596
DB 1152 MLFATSTNLFYLPYEGEISPSNTNIWKNLPRDEIKLIENICWTLMDHLGYPKFMD 1207

RESULT 6
US-10-479-472A-4
; Sequence 4, Application US/10479472A
; Publication No. US2005018581A1
; GENERAL INFORMATION:
; APPLICANT: DEL-FAVERO, JURGEN PETER LODE
; APPLICANT: VAN BROECKHOVEN, CHRISTINE
; TITLE OF INVENTION: NOVEL BRAIN EXPRESSED GENE AND PROTEIN ASSOCIATED WITH
; TITLE OF INVENTION: BIPOLAR DISORDER
; FILE REFERENCE: JAB-1711
; CURRENT APPLICATION NUMBER: US/10/479,472A
; CURRENT FILING DATE: 2003-12-01
; PRIOR APPLICATION NUMBER: PCT/EP02/06316
; PRIOR FILING DATE: 2002-06-06
; PRIOR APPLICATION NUMBER: EP 01202214.1
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin Ver. 3.2
; SEQ ID NO 4
; LENGTH: 1207
; TYPE: PRT
```

```
; ORGANISM: Mus sp.
US-10-479-472A-4

Query Match          90.1%; Score 2835; DB 17; Length 1207;
Best Local Similarity 88.6%; Pred. No. 1.3e-264;
Matches 528; Conservative 36; Mismatches 32; Indels 0; Gaps 0;

QY 1 MNRNGAMDMVDADHYKMFDFDHGNSPMASIOEAEQAABFKKRWTFQVNVVTFQMESTIT 60
DB 612 MNRNGAMDMVDADHYKMFDFDHGNSPMVANOIEAEQAABFKKRWTFQVNVVTFHMESTIT 671

QY 61 RIAYVFGPYNVSSCRFDIDSSNPGIQISLVNNTTEHVSVIVTDYHNLKTRFNYLGGGFF 120
DB 672 RIAYVFGPYNVSSCRFDIDSSSGLQISLVNSTEHSVSVVTDYQNLKSRFSYLGFGGF 731

QY 121 ASVADOGITRFGIGTQAIKVPVRHDIRIIPFGFKFNIAVGLICISLVILTFQWRPYLS 180
DB 732 ASVANOQITRFGIGTQEIIVNVRHDKVNPFPFGFKFNIAVGFILCISLVILTFQWRPYLS 791

QY 181 FRKLMRWILVIALWFIELLDVWSTCSQPICAKWTTEAGSKSLSSSEGHMDLPDVV 240
DB 792 FRKLMRCVLIVIALWFIELLDVWSTCTQPICAKWTTEAKANEKVMISSEGHVVDLPNVI 851

QY 241 ITSLPGSGAEILKQLFNSDDFLYIRVPTAYIDIPETELEDISFVDACEWKVSDIRSGHF 300
DB 852 ITSLPGSGAEILKQLFNSDDFLYIRIPTAYMDIPETELEDISFVDACEWKVSDIRSGHF 911

QY 301 RLRLGWLQSLVQDTKLHLQNIHLHFNPKLAQYFAMNKDKKFKRRESLPSQSRQMGK 360
DB 912 HLLRGWLQSLVQDTKLHLQNIHLHETSRKLAQYFTTNKDKKFKRRESLQDQSRKIG 971

QY 361 AFDRDAEYIRALRRHLVYPSARPVLSSSGSWTLKLHFFQEVLGASMRALYIVRDPRAW 420
DB 972 PFDRDAEYIRALRRHLVYPSARPVLSSSGSWTLKLHFFQEVLGTSMRALYIVRDPRAW 1031

QY 421 IYSMLYNSKPSLSYSLKNVPEHLAKLFKIEGGKGCNLSGYAFEPYELPKELSKSNAV 480
DB 1032 IYSVLGSKPSLSYSLKNVPEHLAKLFKIEGSKCNSNSGYAFEYESLKELETSQSNAI 1091

QY 481 SLLSHLWLANTAALRINTDLPSTYQLVKFEDIIVHPQKTTERRIFAFGLGIPLSPASLNQ 540
DB 1092 SLLSHLWVANTAALRINTDLPNTYHLVKFEDIVHPQKTTERRIFAFGLGIPLSPASLNQ 1151

QY 541 ILFATSTNLFYLPYEGEISPTNTNVKQNLPRDEIKLIENICWTLMDRLGYPKFMD 596
DB 1152 MLFATSTNLFYLPYEGEISPSNTNIWKNLPRDEIKLIENICWTLMDHLGYPKFMD 1207

RESULT 7
US-09-984-429-71
; Sequence 71, Application US/09984429
; Publication No. US20040010132A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 53 Human Secreted Proteins
; FILE REFERENCE: P2018P2
; CURRENT APPLICATION NUMBER: US/09/984,429
; CURRENT FILING DATE: 2001-10-30
; PRIOR APPLICATION NUMBER: 60/244,591
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: 09/288,143
; PRIOR FILING DATE: 1999-04-08
; PRIOR APPLICATION NUMBER: PCT/US98/21142
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 60/061,463
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 60/061,529
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 60/071,498
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 60/061,527
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 60/061,536
```

; PRIOR FILING DATE: 1997-10-09
 ; PRIOR APPLICATION NUMBER: 60/061,532
 ; PRIOR FILING DATE: 1997-10-09
 ; NUMBER OF SEQ ID NOS: 727
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 71
 ; LENGTH: 125
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-984-429-71

Query Match 18.0%; Score 568; DB 11; Length 125;
 Best Local Similarity 100.0%; Pred. No. 1.6e-46;
 Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 185 MRWLLIVIALWFTLELDVWSTCSQPCAKWTRTEAGSKSLSEGHMDLPDVWITSL 244
 DB 1 MRWLLIVIALWFTLELDVWSTCSQPCAKWTRTEAGSKSLSEGHMDLPDVWITSL 60
 QY 245 PGSGAEILKQLFFNSSDFLYIRVPTAYIDIPETELEIDSFVDACEWK 291
 DB 61 PGSGAEILKQLFFNSSDFLYIRVPTAYIDIPETELEIDSFVDACEWK 107

RESULT 8
 US-10-150-111-71
 ; Sequence 71, Application US/10150111
 ; Publication No. US20030078386A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rubin et al.
 ; TITLE OF INVENTION: Secreted Protein HPEAD48
 ; FILE REFERENCE: P2018PDI
 ; CURRENT APPLICATION NUMBER: US/10/150.111
 ; CURRENT FILING DATE: 2002-05-20
 ; PRIOR APPLICATION NUMBER: 09/288,143
 ; PRIOR FILING DATE: 1999-04-08
 ; PRIOR APPLICATION NUMBER: PCT/US98/21142
 ; PRIOR FILING DATE: 1998-10-08
 ; PRIOR APPLICATION NUMBER: 60/061,463
 ; PRIOR FILING DATE: 1997-10-09
 ; PRIOR APPLICATION NUMBER: 60/061,529
 ; PRIOR FILING DATE: 1997-10-09
 ; PRIOR APPLICATION NUMBER: 60/071,498
 ; PRIOR FILING DATE: 1997-10-09
 ; PRIOR APPLICATION NUMBER: 60/061,527
 ; PRIOR FILING DATE: 1997-10-09
 ; PRIOR APPLICATION NUMBER: 60/061,536
 ; PRIOR FILING DATE: 1997-10-09
 ; PRIOR APPLICATION NUMBER: 60/061,532
 ; PRIOR FILING DATE: 1997-10-09
 ; NUMBER OF SEQ ID NOS: 219
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 71
 ; LENGTH: 126
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: SITE
 ; LOCATION: (126)
 ; OTHER INFORMATION: Xaa equals stop translation
 US-10-150-111-71

Query Match 18.0%; Score 568; DB 14; Length 126;
 Best Local Similarity 100.0%; Pred. No. 1.6e-46;
 Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 185 MRWLLIVIALWFTLELDVWSTCSQPCAKWTRTEAGSKSLSEGHMDLPDVWITSL 244
 DB 1 MRWLLIVIALWFTLELDVWSTCSQPCAKWTRTEAGSKSLSEGHMDLPDVWITSL 60
 QY 245 PGSGAEILKQLFFNSSDFLYIRVPTAYIDIPETELEIDSFVDACEWK 291
 DB 61 PGSGAEILKQLFFNSSDFLYIRVPTAYIDIPETELEIDSFVDACEWK 107

RESULT 9
 US-10-021-660-128
 ; Sequence 128, Application US/10021660
 ; Publication No. US20030152926A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Murray, Richard
 ; APPLICANT: Glynn, Richard
 ; APPLICANT: Watson, Susan R.
 ; APPLICANT: EOS Biotechnology, Inc.
 ; TITLE OF INVENTION: No. US20030152926A1el Methods of Diagnosis of Angiogenesis,
 ; TITLE OF INVENTION: Compositions and Methods of Screening for Angiogenesis
 ; TITLE OF INVENTION: Modulators
 ; FILE REFERENCE: 018501-000710US
 ; CURRENT APPLICATION NUMBER: US/10/021,660
 ; CURRENT FILING DATE: 2001-12-06
 ; PRIOR APPLICATION NUMBER: US/09/784,356
 ; PRIOR FILING DATE: 2001-02-14
 ; PRIOR APPLICATION NUMBER: US 09/637,977
 ; PRIOR FILING DATE: 2000-08-11
 ; NUMBER OF SEQ ID NOS: 135
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 128
 ; LENGTH: 411
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-021-660-128

Query Match 4.7%; Score 147; DB 14; Length 411;
 Best Local Similarity 18.2%; Pred. No. 6.4e-05;
 Matches 89; Conservative 58; Mismatches 149; Indels 194; Gaps 16;
 QY 187 WLLIVIALWFTLELDVWSTCSQPCAKWTRTEAGSKSLSS 229
 DB 5 WKAVLLALASIAIQYTAITFTAKSPHTCPGLAAGLAERLC-----EESPTAYNLSR 59
 QY 230 EGHMDLPDVWITSLPGSGAEILKQLFFNSSDFLYIRVPTAYIDIPETELEIDSFVDACE 289
 DB 60 KTH-----ILILATTSRSGSFVQLFNQHLDPVLFLEPL----- 93
 QY 290 WKVSDIRSGHFRLLRGLWQLSLVQDTKLHLQNIHLHEPNRGLAQYFAMNKDKKPKRRE 349
 DB 94 -----YHVQNTLIPFTQG-----KSPADRRV 115
 QY 350 SLPEQRSMKGARDRAEYI-----RALRRHLVYVPSRPV-----LS 387
 DB 116 MLGASRDLLSLYDCDLYFLENYIKPPVNHHTDRIFRGASRVLCSPVCDPPGADLV 175
 QY 388 LSSGSWTLKLHFFQEVILGA-----SMRALYIVRD 416
 DB 176 LEEGDCVRKCGLLNLTVAAEACRSHVAIKTVRVPVNDLRALVEDPRNLKVIQLVRD 235
 QY 417 PRAWIYMLYNKPSLYSLKNVPEHLAKPKIEGKGKCNLNSGYAFEPYELRKLKSKK 476
 DB 236 PGILASRSETFRDTY-----RLWRLWYGTGR-----KYNLDVTQL-TTVCSDF 279
 QY 477 SNAVS--LLSHLWLANTAALRINTDLLPTSQYQVKEEDIVHFPQKTERIFAFIGPL- 533
 DB 280 SNSVSTGLMRPPMLKG-----KYMVLVRYVEDLARNPMKTEEYIGLIGLIPD 325
 QY 534 -----SPASLNQILFATSTNLFVLPYEGEISPTNTNWNKQNLPRDEIKLIENI 581
 DB 326 SHVARWTONNTRGDPITLKGKHYGTVRN-----SAATAEKWRFLSLDIVAFAGNA 375
 QY 582 CWTLMRLGY 591
 DB 376 CQVLAQLGY 385

RESULT 10
 US-10-211-462-97
 ; Sequence 97, Application US/10211462

```
; Publication No. US20040033495A1
; GENERAL INFORMATION:
; APPLICANT: Murray, Richard
; APPLICANT: Glynn, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Aziz, Natasha
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Angiogenesis, Compositions and
; TITLE OF INVENTION: Methods of Screening for Angiogenesis Modulators
; FILE REFERENCE: 018501-006200US
; CURRENT APPLICATION NUMBER: US/10/211,462
; CURRENT FILING DATE: 2003-02-13
; PRIOR APPLICATION NUMBER: US 09/784,356
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: US 09/791,390
; PRIOR FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: US 60/310,025
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US 60/334,244
; PRIOR FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 230
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 97
; LENGTH: 411
; TYPE: PR1
; ORGANISM: Homo sapiens
US-10-211-462-97
```

```
Query Match 4.7%; Score 147; DB 15; Length 411;
Best Local Similarity 18.2%; Pred. No. 6.4e-05;
Matches 89; Conservative 58; Mismatches 149; Indels 194; Gaps 16;

QY 187 WILVIALWFIEL-----LDVWSTC-----SQPICKWTRTEAGSKSLSS 229
Db 5 WKAVLLALASIAIQYTAIRTTAKSFHTCPGLAELAEALRC-----ESPTFAYNLSR 59

QY 230 EGHMDLPDVVITSLPGSGAEILKQLFFNSDFLYIRVPTAYIDIPETELEIDSFVDACE 289
Db 60 KTH-----ILLATRSGSGFVQGFNQHLDFVYLFPEL----- 93

QY 290 WKVSDIRSGHFRLLRGWLQSLVQDTKLHLQNIHLHFNPNRCKLAQYFAMNKDKRKFQRE 349
Db 94 -----YHVQNTLIPRFTQG-----KSPADRRV 115

QY 350 SLPEQRSMKGAFDRDAEYI-----RALRRHLVYPSARPV-----LS 387
Db 116 MLAGSRDLRLSLYDCDLYFLENYKPPPVNHTTDRIFRFGASRVLCSPVCDPPGPADLV 175

QY 388 LSSGSWTLKLHFFQEVLGA-----SMRALYIVRD 416
Db 176 LEEGDCVRKCGLLNLTVAAEACRERSHVAIKTVRVPVNDLRALVEDPRLNLKVIQLVRD 235

QY 417 PRAWIYMLYNSKPSLYSLKNVPEHLAKLFKEGGKGNLNSGYAFEYELPKELSKSK 476
Db 236 PRGILASRSETFRDTY-----RLWRLWYGTGR-----KPYNLDTVQL-TTVCEDF 279

QY 477 SNAVS--LLSHLWLANATAALRINTDLLPTSQYLVKFEIDIVHPQKTTIRIFAFGLIPL- 533
Db 280 SNSVSTGLMRPPWLKG-----KMYLVRYEDLARNPMKTEIYFGLGIPLD 325

QY 534 -----SPASLNQILFATSTNLFYLPVEGEISPTNTNVWKNLPRDIKLIENI 581
Db 326 SHVARWQNTRGDPITLKGKHYGTVRN-----SAATAEKWRFLSYDIVAFAQNA 375

QY 582 CWTLMRDLGY 591
Db 376 CQOVLAQLGY 385
```

RESULT 11

```
US-10-408-765A-395
; Sequence 395, Application US/10408765A
; Publication No. US20040101874A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Rahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 395
; LENGTH: 411
; TYPE: PR1
; ORGANISM: Homo sapiens
US-10-408-765A-395
```

```
Query Match 4.7%; Score 147; DB 16; Length 411;
Best Local Similarity 18.2%; Pred. No. 6.4e-05;
Matches 89; Conservative 58; Mismatches 149; Indels 194; Gaps 16;

QY 187 WILVIALWFIEL-----LDVWSTC-----SQPICKWTRTEAGSKSLSS 229
Db 5 WKAVLLALASIAIQYTAIRTTAKSFHTCPGLAELAEALRC-----ESPTFAYNLSR 59

QY 230 EGHMDLPDVVITSLPGSGAEILKQLFFNSDFLYIRVPTAYIDIPETELEIDSFVDACE 289
Db 60 KTH-----ILLATRSGSGFVQGFNQHLDFVYLFPEL----- 93

QY 290 WKVSDIRSGHFRLLRGWLQSLVQDTKLHLQNIHLHFNPNRCKLAQYFAMNKDKRKFQRE 349
Db 94 -----YHVQNTLIPRFTQG-----KSPADRRV 115

QY 350 SLPEQRSMKGAFDRDAEYI-----RALRRHLVYPSARPV-----LS 387
Db 116 MLAGSRDLRLSLYDCDLYFLENYKPPPVNHTTDRIFRFGASRVLCSPVCDPPGPADLV 175

QY 388 LSSGSWTLKLHFFQEVLGA-----SMRALYIVRD 416
Db 176 LEEGDCVRKCGLLNLTVAAEACRERSHVAIKTVRVPVNDLRALVEDPRLNLKVIQLVRD 235

QY 417 PRAWIYMLYNSKPSLYSLKNVPEHLAKLFKEGGKGNLNSGYAFEYELPKELSKSK 476
Db 236 PRGILASRSETFRDTY-----RLWRLWYGTGR-----KPYNLDTVQL-TTVCEDF 279

QY 477 SNAVS--LLSHLWLANATAALRINTDLLPTSQYLVKFEIDIVHPQKTTIRIFAFGLIPL- 533
Db 280 SNSVSTGLMRPPWLKG-----KMYLVRYEDLARNPMKTEIYFGLGIPLD 325

QY 534 -----SPASLNQILFATSTNLFYLPVEGEISPTNTNVWKNLPRDIKLIENI 581
Db 326 SHVARWQNTRGDPITLKGKHYGTVRN-----SAATAEKWRFLSYDIVAFAQNA 375

QY 582 CWTLMRDLGY 591
Db 376 CQOVLAQLGY 385
```

RESULT 12

```
US-10-723-860-1544
; Sequence 1544, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882.0193.NPUS01
```

```

RESULT 13
US-10-212-933-2
; Sequence 2, Application US/10212933
; Publication No. US20030008366A1
; GENERAL INFORMATION:
; APPLICANT: Uchimura, Kenji
; APPLICANT: Muramatsu, Hideki
; APPLICANT: Kadomatsu, Kenji
; APPLICANT: Kannagi, Reiji
; APPLICANT: Habuchi, Osami
; APPLICANT: Muramatsu, Takeshi
; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGUCOSAMINE-6-0-
; TITLE OF INVENTION: SULFOTRANSFERASE AND
; TITLE OF INVENTION: DNA ENCODING THE SAME
; FILE REFERENCE: TOYAM41.001AUS
; CURRENT APPLICATION NUMBER: US/10/212,933
; CURRENT FILING DATE: 2002-08-05
; PRIOR APPLICATION NUMBER: US/09/263,023
; PRIOR FILING DATE: 1999-03-05
; PRIOR APPLICATION NUMBER: JP 10-54007

```

Db 338 GGPADYHAL-----GAMEVICNSMAKTLQTLQ-PPDWLQGHYLVVRYEDLVGDPV 387
Qy 520 KTTTERIPAFILGIPSPA-----SLNQILFATSTNLFLPYEGEISPTN----TNVWKQNL P 571
Db 388 KTLRRVYDFVGLLVSPMEQFALNMTSGSGSSKPFV-----VSARNATQAANAWRTALT 442
Qy 572 RDEIKLIENICWTIMDRLGYPK 593
Db 443 FQIKQVEEFCYOPMAVLGYER 464

RESULT 15
US-10-723-860-1409
; Sequence 1409, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1409
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-723-860-1409

Query Match 4.2%; Score 133; DB 16; Length 530;
Best Local Similarity 26.2%; Pred. No. 0.0021;
Matches 53; Conservative 36; Mismatches 81; Indels 32; Gaps 9;

Qy 408 MRALYIVRDPRAITYSLYNSKPSLY--SLKNYPE-----HLAKLPKIEGK-GKCNLNS 459
Db 325 LKVIHLVRDPRV-VASSRIRSHGLIRESLQVVRSDPRRAHRMPFLEAAGHKLGAKEGV 383
Qy 460 GYAFEYELRKLKLSKSNVSLSHLWLANATAALRINTDLLPTSYQLVKFEDI VHF PQ 519
Db 384 GGPADYHAL-----GAMEVICNSMAKTLQTLQ-PPDWLQGHYLVVRYEDLVGDPV 433
Qy 520 KTTTERIPAFILGIPSPA-----SLNQILFATSTNLFLPYEGEISPTN----TNVWKQNL P 571
Db 434 KTLRRVYDFVGLLVSPMEQFALNMTSGSGSSKPFV-----VSARNATQAANAWRTALT 488
Qy 572 RDEIKLIENICWTIMDRLGYPK 593
Db 489 FQIKQVEEFCYOPMAVLGYER 510

Search completed: June 23, 2005, 09:27:55
Job time : 77.3449 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2005, 08:33:38 ; Search time 13.5183 Seconds
(without alignments)
4242.027 Million cell updates/sec

Title: US-10-697-828-9
Perfect score: 3147
Sequence: 1 MNRNGAMDMVDVAHYKMF.....LIENICWTLMMDRLGYPKFM 596

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_79:.*
1: Pir1.*
2: Pir2.*
3: Pir3.*
4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|--------|--------------------|
| 1 | 133 | 4.2 | 484 | JE0261 | N-acetylglucosamin |
| 2 | 113.5 | 3.6 | 484 | JC7350 | N-acetylglucosamin |
| 3 | 109.5 | 3.5 | 1441 | T00335 | hypothetical prote |
| 4 | 107 | 3.4 | 2401 | T28676 | rhostry protein - |
| 5 | 106.5 | 3.4 | 532 | H96840 | hypothetical prote |
| 6 | 106 | 3.4 | 486 | JC7351 | N-acetylglucosamin |
| 7 | 105.5 | 3.4 | 518 | B83472 | hypothetical prote |
| 8 | 103 | 3.3 | 582 | T46822 | phytoene desaturas |
| 9 | 102.5 | 3.3 | 689 | S45901 | probable membrane |
| 10 | 102 | 3.2 | 1333 | S65812 | DNA-directed DNA p |
| 11 | 101.5 | 3.2 | 591 | B97994 | RNA primase (EC 2. |
| 12 | 101 | 3.2 | 486 | H81377 | exopolysphatase |
| 13 | 100 | 3.2 | 1663 | T28923 | hypothetical prote |
| 14 | 98 | 3.1 | 480 | T20168 | hypothetical prote |
| 15 | 98 | 3.1 | 586 | H95123 | DNA primase (limpr |
| 16 | 97.5 | 3.1 | 834 | T28250 | ORF MSV089 probabl |
| 17 | 97 | 3.1 | 420 | A28882 | alpha-1-antitrypsi |
| 18 | 97 | 3.1 | 561 | G54246 | conserved hypothet |
| 19 | 96 | 3.1 | 458 | A57397 | chondroitin 6-sulf |
| 20 | 96 | 3.1 | 1692 | G01449 | probable mucin G2 |
| 21 | 96 | 3.1 | 4910 | S64942 | probable membrane |
| 22 | 95 | 3.0 | 522 | B86294 | hypothetical prote |
| 23 | 95 | 3.0 | 963 | C90535 | conserved hypothet |
| 24 | 94 | 3.0 | 1405 | T27969 | hypothetical prote |
| 25 | 93.5 | 3.0 | 804 | A39972 | segment S5 protein |
| 26 | 93.5 | 3.0 | 825 | T32130 | hypothetical prote |
| 27 | 93 | 3.0 | 648 | S56783 | hypothetical prote |
| 28 | 93 | 3.0 | 676 | A91349 | probable ATP-depen |
| 29 | 93 | 3.0 | 868 | S65186 | NIP80 protein - ye |

conserved hypothet
ABC transport syst
hypothetical prote
hypothetical prote
dynein heavy chain
hypothetical prote
hypothetical prote
probable receptor-
probable 2-acylgly
probable CDP glyce
ATP-dependent heli
hypothetical prote
hypothetical prote
limb deformity (ld
hypothetical prote

ALIGNMENTS

RESULT 1

JE0261
N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) - human
C:Species: Homo sapiens (man)
C:Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 01-Mar-2002
C:Accession: JE0261
R:Uchimura, K.; Muramatsu, H.; Kaname, T.; Ogawa, H.; Yamakawa, T.; Fan, Q.W.; Mitsuoka, T.

J. Biochem. 124, 670-678, 1998

A:Title: Human N-Acetylglucosamine-6-O-sulfotransferase Involved in the biosynthesis of (

A:Reference number: JE0261; MUID:98391845; PMID:9722682

A:Accession: JE0261

A:Molecule type: mRNA

A:Residues: 1-484 <UCH>

A:Cross-references: DDBJ:AB014679

C:Comment: This protein catalyzes the transfer of sulfer of sulfate from 3'-phosphoadenos

C:Superfamily: chondroitin 6-sulfotransferase

C:Keywords: sulfotransferase

Query Match 4.2%; Score 133; DB 2; Length 484;

Best Local Similarity 26.3%; Pred. No. 0.023;

Matches 53; Conservative 36; Mismatches 81; Indels 32; Gaps 9;

| | | | | |
|----|-----|---|----------------------|-----|
| Qy | 408 | MRALYIVRDPRAWYMLYNSKPSLY--SLKNVPE---- | HLAKLPKIEGSK-GKCNLNS | 459 |
| Db | 279 | LKVIHLVRDPRA-VASSRIRSRHGLIRESLQVRSRDPRAHRMPFLEAAGHLGAKKEGV | 337 | |
| Qy | 460 | GVAPEYEPRLKELSKSNVAVLSLHLWLAATAALINTDLLPTSQVLKVFEDIVHPFQ | 519 | |
| Db | 338 | GGPADYHAL-----GAMEVICNSMAKTLQTALQ-PPDWLQGHVLYVRVEDLVGEPV | 387 | |
| Qy | 520 | KTERIPAFIGIPLSPA-----SLNQILFATSTNLFYLPYEGETSPTN-----TNVWKQLNP | 571 | |
| Db | 388 | KTLRRVYDFVGLLVSPMEQFALNMTSGSGSSKPFV-----VSARNATQAANARALT | 442 | |
| Qy | 572 | RDEIKLIENICWTLMDBLGYPK | 593 | |
| Db | 443 | FQOIQVQEEFCYQPMVILGYER | 464 | |

RESULT 2

JC7350
N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) [validated] - mouse

C:Species: Mus musculus (house mouse)

C:Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 09-Jul-2004

C:Accession: JC7350

R:Uchimura, K.; Fasakhany, F.; Kadamatsu, K.; Matsuoka, T.; Yamakawa, T.; Kurosawa, N.;

Biochem. Biophys. Res. Commun. 274, 291-296, 2000

A:Title: Diversity of N-Acetylglucosamine-6-O-sulfotransferases: Molecular cloning of a r

A:Reference number: JC7350; MUID:20374462; PMID:10913333

A:Accession: JC7350


```

Db      989  LNKIKIIDVKYBTDEIEKNKKNKINAELSSEKIITQLKENSLSKQCQSIKSTIDDNY 1048
Qy      366  ---AEVIRALRRHLV-----YPSARPV---LSLSSGSLTKLHLFFQEVLGSMRA 410
Db      1049  VSECIKNITLKVIVNEKKNINITYFNABEYNQVSLN-----FNNIEMADTKS 1098
Qy      411  LYIVRPRAWIYMLYNSKPSLSKKNVPEHLAK--LFKIEGGKG---KCN-----456
Db      1099  QYILNKKCN-----NGTNTDYNIKELKEHKKSNVYKDEAGKNTQBIKKNKELFEKYE 1152
Qy      457  -----LNSGVAFPEYELRKLKSKSNVALLSHLWLAN---TAAALRINTDLPTSQY 507
Db      1153  QEVTVLLKKYAVE---LKNKFDKTKYSEQIIKEIKDAHNTFTSQADKSKKQNEIKNE 1209
Qy      508  LKVEDIVHPQKTERIFAFGLGTPLASLNQILFATSTNLFPYLPVE--GEISPTNTV 565
Db      1210  QIRIEDEVAKNNKNK---AILDQLSVEFP-KIKFLKIKDLRTKSDCLKETKDIETKI 1265
Qy      566  WKQNLPRDEIKLIEN 580
Db      1266  SNLSIDTQETKLIEN 1280

RESULT 5
H96840
hypothetical protein F23A5.18 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: H96840
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federpspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-532 <STO>
A:Cross-references: UNIPROT:Q9SAH8; GB:AE005173; NID:G6503294; PIDN:AAF14670.1; GSPDB:GN
C:Genetics:
A:Gene: F23A5.18
A:Map position: 1
C:Superfamily: natural resistance-associated macrophage protein 1

Query Match      3.4%; Score 106.5; DB 2; Length 532;
Best Local Similarity 19.7%; Pred. No. 2.9;
Matches 125; Conservative 90; Mismatches 210; Indels 209; Gaps 33;

Qy      26  NSPMASIOEAEQ-AAEFKKRTQFNVVTFQMESTITRIAYVYGPY-----70
Db      23  NSPLIENSDSQIIVSEKSKWNF--FAYLPGFLVLSIAYIDPGNFETDLQAGAHYKEL 80
Qy      71  ---INVSSCRFTDSSNPGQLQISLVNNTNTHVVSIVTDYHNLKTRFNLYLFGFGFASVADQG 127
Db      81  LWIILVASC-----AALVQSLAANLGVVTKHLAE-----QCRAEYS 118
Qy      128  QITRPLGTQAIKVPVRHD--RII---PPFGKFNIAV--GLIL--CISLVILTFQWRFY 178
Db      119  KVPNFMWVAEIAVACCDIPEVIGTAFALNMLFSPVWIGVLLTGLSTLILLALQKYGV 178
Qy      179  LSFRLKMRWILLVIALWFIELLDVWSTCSQPICAKWTTRTEAE--GSKKSLSEGHMDLP 238
Db      179  RKLEFLIAFLVPTTIAICFFVEL-----HYSPD 206
Qy      239  -----VVITSLPGSGAEIL-----KQLFFNSDFLYIRVPTAVIDIPETELE 280

```

```

Db      207  PGEVLHGLFVPQIKNGCATGLAISLIGAMVPHNLFLHSALVLSRKIPRSGSIKE-----262
Qy      281  IDSFPDACEKVKSDIRSGHFRLLRGLWLSQVQDTKHLQNIHLHHPNRGKLAQYFAMNKD 340
Db      263  -----ACRFYL--IESG---LALMVAFLINVSIVSGAVCNAPNLS-----299
Qy      341  KKRKFKRRESLPRQSOMKCAFDRDAEYIRALRRHLVYVPSARPVLSLSSGSLTKLHFF 400
Db      300  -----PEDRANCE---DLDLNKASFLLRNVV-----GKWSKUL--F 330
Qy      401  QEVLGASMR-----LYIVR---DPR--AWIYSMLYNSKPSLSKKNVPEHLAKLF 446
Db      331  AIALLAGCQSTTITGYAGQYVMQGLDLRLPEWLRNLTR-----CLAIPSLIVALI 384
Qy      447  KIEGGKGCNLSNGYAFEYE-----PLRKELS-KSKS-----NAVSLSLHLWLANAAL 495
Db      385  GGSAGAGKLIIIASMILSPALVELLFTCKTKGSHVNPMAITALTWTWIG-GLIM 443
Qy      496  RINTDLLPTSQYLVKVEDIVHFPQKTERIF-----AFLGIPLSASLNQILF---ATSTN 548
Db      444  GINIYVLVSF--IKL--LIHSHMKLILVVFCGILGSPAGIALYLAAYLVRKRNVRATS 499
Qy      549  LFYLPYEGEISPTNTNWKQNLPRDEIKLIENIC 582
Db      500  LL-----ISRDSQNV---ETLPRQDIVNMQLPC 524

RESULT 6
JC7351
N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) [validated] - human
C:Species: Homo sapiens (man)
C>Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 09-Jul-2004
C:Accession: JC7351
R:Uchimura, K.; Pasakany, F.; Kadomatsu, K.; Matsukawa, T.; Yamakawa, N.;
Biochem. Biophys. Res. Commun. 274, 291-296, 2000
A:Title: Diversity of N-acetylglucosamine-6-O-sulfotransferases: Molecular cloning of a
A:Reference number: JC7350; MUID:20374462; PMID:10913333
A:Accession: JC7351
A:Molecule type: mRNA
A:Residues: 1-486 <UCH>
A:Cross-references: UNIPROT:O75667; DDBJ:AB040711
C:Comment: This enzyme, having a type II transmembrane topology and N-linked glycosylatic
sulfate linkage formation. This enzyme is involved in synthesis of L-selectin ligand in
C:Keywords: glycolysis; sulfate transport; sulfotransferase; transmembrane protein

Query Match      3.4%; Score 106; DB 2; Length 486;
Best Local Similarity 20.6%; Pred. No. 2.8;
Matches 87; Conservative 45; Mismatches 136; Indels 154; Gaps 18;

Qy      179  LSFRLKMRWILLVIALWFIELLDVWSTCSQPICAKWTTRTEAE--GSKKSLSEGHMDLP 237
Db      187  LTTAALFRWTKVI-----CSPPLCPGAPARAEEVLVEDTACE---RSCP 230
Qy      238  DVITSLPGSGAEILKOLFNSDFLYIRVPTAVIDIPETELEIDSFVDACEWKVSDIRS 297
Db      231  PVAIRALE---AECKR-----YPVVVI-----KQVR- 253
Qy      298  GHFLLRGLWLSQVQDTKHLQNIHLHHPNRGKLAQYFAMNKDKRFRRESLPEORSQ 357
Db      254  ---LLDLGLVPLLRDPGLNLKVVQLFRDPRA-----VHNSRLKSRQGLLRSTQLVLR 305
Qy      358  MKGAFDRDAEYIRALRRHLVYVPSARPVLSLSSGSLTKLHLFFQEVLGASMRALYIVRDP 417
Db      306  QRG--DR---FHRVLLAHGV---GARP-----GGQSRALPAA--P 335
Qy      418  RAWIYSMLYNSKPSLSKKNVPEHLAKLFKTEGGKGCNLSNGYAFEYELRKLKSKS 477
Db      336  RADFF-----LT 342
Qy      478  NAVSLSLHLWLANAALRINTDLPTSQYLVKVEDIVHFPQKTERIFAFGLI-----PL 533
Db      343  GALEVICAWLRDLFLA-RGAPAWLRRYRLRYREDLVRQRAQLRRLRSLGRLAAL 401

```


Qy 577 L---IENICWTAMD 587
| | | | |
Db 840 LSHPEHRTKLTVD 853

RESULT 14

T20168
hypothetical protein C53A5.9 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T20168
R:Mortimore, B.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19232
A:Accession: T20168
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-480 <WIL>
A:Cross-references: UNIPROT:O17700; EMBL:Z81486; PIDN:CAB03989.1; GSPDB:GN00023; CESP:CS53A5
A:Experimental source: clone C53A5
C:Genetics:
A:Gene: CESP:C53A5.9
A:Map position: 5
A:Introns: 28/1; 59/3; 194/1; 264/3; 332/1

| | | | | |
|-----------------------|-------|--------------|------|--------------------------------------|
| Query Match | 3.1% | Score 98 | DB 2 | Length 480 |
| Best Local Similarity | 20.6% | Pred. No. 12 | | |
| Matches | 94 | Conservative | 60 | Mismatches 162; Indels 140; Gaps 22; |

Qy 199 ELLDVWSTCSQPI---CAKWTREAGSKSLSEGHMDLPDVVITSLPGSGAEILKQL 255
| | | | |
Db 67 EVADAWDCQPLEYQDAQFSESEMETQMKEL--EDCSIEASRVFLENLQKS-FKLAKLL 123

Qy 256 ----FFNSDPLYRVPTAYIDIPETLEIDSFVDACEKWKSDIRSGHFRLLRWGLQSLV 311
| | | | |
Db 124 QLTQSLNPQESSYNR-----KIISTEMK-----KINDTVS-----QLKY 157

Qy 312 QDTKLHLQNTLHL-----ENRCKLA---QYFAMNK-----DKRKFPR 347
| | | | |
Db 158 AKRIEMINEHLEKHPFGKDGDAFLBNPMDPVLIAOQAQAHKPLLIGSDPDKTKYP 217

Qy 348 RESLPEQSQMGAFDPAEYIRALRRHLVYPSARPVLSSSGSWTLKLHFFOE----- 402
| | | | |
Db 218 IGRLPDPANFAVASSKTNLV-----IGTHNGQFLNKFYNNQKNAR 262

Qy 403 VLGSAMR-----ALYIVRDPRAWIYMLNYSKPSLYSL-----KNVPE----- 440
| | | | |
Db 263 CMGANLNHKRTKPAAGFHNGALYVVG-----YDSIVLSSVELYDLBEGTWKGPSLNNC 317

Qy 441 -----HLAKLFKEGGKCKNLN-----SGYAFEPYELRKLKSLSKSNVSLLSH 485
| | | | |
Db 318 RANAIVACYGEIFVLGFGNGKNEESIEKISAGNEFEIFG-EMEGRSFGACVQGR 376

Qy 486 LWLA-----NTAAALRLNDLLPT-----SYQLVKPEDIVHPQKTKTERIFAFGLIPL 533
| | | | |
Db 377 IYIAGSNTNTLTKSVRSYDPYTKTWDEFSMKNAKAFILH--ATNEALYAIRGYDE 433

Qy 534 SPASLNQILFATSTNLFY--LPYGEISPTNTNVWK 567
| | | | |
Db 434 ESALLDQIERDPKKLKWSIVPSKPHVPPTSYNYEK 469

RESULT 15

H95123
DNA primase [imported] - *Streptococcus pneumoniae* (strain TIGR4)
C:Species: *Streptococcus pneumoniae*
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 09-Jul-2004
C:Accession: H95123
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Maynam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001

A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete Genome Sequence of a virulent isolate of *Streptococcus pneumoniae*.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: H95123
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-586 <KUR>
A:Cross-references: UNIPROT:Q97QX2; GB:AE005672; PIDN:AAK75185.1; PID:gl4972547; GSPDB:GB
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SPI072

| | | | | |
|-----------------------|-------|--------------|------|--------------------------------------|
| Query Match | 3.1% | Score 98 | DB 2 | Length 586 |
| Best Local Similarity | 18.8% | Pred. No. 15 | | |
| Matches | 117 | Conservative | 83 | Mismatches 212; Indels 210; Gaps 27; |

Qy 87 QISLVNNTTEHVVSIVTDYHNL-KTRNYLGG---FGG-----PASVADQGGQITRFLGLTQ 137
| | | | |
Db 5 QVIEEIKNNANIVVEIGDVLSLQAGRNYYLQCPFHGKPTSFNVNVEDKQFYHCFGGGRS 64

Qy 138 AIVKPVHRHRIIPFGFGKFNIAVGLILCISLVILTQWRFYLSFRKLMRWLILVIALWPF 197
| | | | |
Db 65 GDV-----PKF-----IEEYQ-----GVPPF 79

Qy 198 IELLDVWSTCSQPICAKWTREAGSKSLS---SEGHMDLPD-----VVITS 243
| | | | |
Db 80 IEAVQILG-----QRVGIEVEKPLYSEQKSPASPHQALYDMHEDAAKFYHAILMTT 129

Qy 244 LPGSGA-----EILKQLFFN-----SSDFLYIRVPTAY--IDIPETEL---BID 282
| | | | |
Db 130 TMGEARNYLYQRLTDEVILKHFHIGLAPPERNLYLQRLSDQYREEDLLDSGLFVLSDN 189

Qy 283 SFVDACE-----WKVSDIRSGHFRLLRWGLQSLVQDTKLHLQNI 321
| | | | |
Db 190 QFVDTFNRMPLTNDQGVIAFSGRIMQKTDSTQSKYNSR---STAIENKSYELVYM 246

Qy 322 HLHPEPNKGLAQYFAMNKKKRPKRRESLPEQBSQMKGAPDRD-AEYIRALRRHLVYYP 380
| | | | |
Db 247 DRAKRSSGKASEIYLMEGFMDVIAAYRAGIENAVASMGTALSREHVEHLKLTKKLV--- 303

Qy 381 SARPVLSGSGSWTLKLHFPQEVLG--ASMRALYIVRDPRAWIYMLNYSKPSLYSLQNV 438
| | | | |
Db 304 -----LVYDGDKAGQAATLKDDEIGDMPVQIVSMPPDNLDPDEYLQNG 347

Qy 439 PEHLAKLFKEGGKCKNLNSGYAFEYELRKLKSLKSNVSLLSHLWLANTAAA----- 494
| | | | |
Db 348 PEDLAYLLT---KTRISPIEFYIHOYKPNSENLOAQIEPLEKIAPLIVQEKSTIAAQN 403

Qy 495 -LRINTDLLPTSQVLVKFEDIVH-----FPQKTERIFAFGLIPL- 533
| | | | |
Db 404 YIHILADSL-ASFDYTIQIEQVNESRQVRQNRMEGISRPTPIIMPVTTKQLSALMRAEAH 462

Qy 534 -----SPASLNQIL-----FATSTNLFYLPYE-----GEISP-----TNTNVWKQN 569
| | | | |
Db 463 LLYRMESPLVNDYRLREDFAFATPEFQVLYDLQYGNLPPEVLAEQBTSEVERAWYQV 522

Qy 570 LPRD-----EIKLIEN 580
| | | | |
Db 523 LAQDLPAEISPOELSEVENTRN 544

Search completed: June 23, 2005, 08:50:15
Job time : 16.5183 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2005, 08:33:13 ; Search time 65.4049 Seconds
(without alignments)
4666.302 Million cell updates/sec

Title: US-10-697-828-9
Perfect score: 3147
Sequence: 1 MNRNGAMMDVDAHYKMF.....LIENICWTLMDRLGYPKFM 596

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_03.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|----------|---------------------|
| 1 | 3142 | 99.8 | 1222 | 2 Q81ZU8 | Q81ZU8 homo sapien |
| 2 | 1821 | 57.9 | 755 | 2 Q9HAK5 | Q9HAK5 homo sapien |
| 3 | 920 | 29.2 | 480 | 2 Q9H951 | Q9H951 homo sapien |
| 4 | 863 | 27.4 | 784 | 2 Q6P523 | Q6P523 homo sapien |
| 5 | 615 | 19.5 | 314 | 2 Q9HAD7 | Q9HAD7 homo sapien |
| 6 | 150 | 4.8 | 411 | 2 Q9EQC0 | Q9EQC0 mus musculus |
| 7 | 147 | 4.7 | 411 | 2 Q43916 | Q43916 homo sapien |
| 8 | 136 | 4.3 | 958 | 2 Q8UL01 | Q8UL01 homo sapien |
| 9 | 136 | 4.3 | 958 | 2 Q8BL14 | Q8BL14 m mus muscu |
| 10 | 134 | 4.3 | 483 | 2 Q794G9 | Q794G9 mus musculus |
| 11 | 134 | 4.3 | 530 | 2 Q88276 | Q88276 mus musculus |
| 12 | 134 | 4.3 | 530 | 2 Q80WV3 | Q80WV3 mus musculus |
| 13 | 133 | 4.2 | 483 | 2 Q9UE05 | Q9UE05 homo sapien |
| 14 | 133 | 4.2 | 530 | 2 Q9V4C5 | Q9V4C5 homo sapien |
| 15 | 129 | 4.1 | 304 | 2 Q677T4 | Q677T4 lymphocyti |
| 16 | 128 | 4.1 | 486 | 2 Q9VMC3 | Q9VMC3 drosophila |
| 17 | 127 | 4.0 | 441 | 2 Q93403 | Q93403 torpedo cal |
| 18 | 126.5 | 4.0 | 420 | 2 Q6DBY9 | Q6DBY9 brachydanio |
| 19 | 126 | 4.0 | 388 | 2 Q9WJES | Q9WJES mus musculus |
| 20 | 125 | 4.0 | 388 | 2 Q9R111 | Q9R111 mus musculus |
| 21 | 123.5 | 3.9 | 395 | 2 Q6GZK3 | Q6GZK3 homo sapien |
| 22 | 122.5 | 3.9 | 384 | 2 Q6ZE88 | Q6ZE88 burkholderi |
| 23 | 122.5 | 3.9 | 479 | 2 Q7LGC8 | Q7LGC8 homo sapien |
| 24 | 122 | 3.9 | 392 | 2 Q6GN39 | Q6GN39 xenopus lae |
| 25 | 122 | 3.9 | 515 | 2 Q8HV31 | Q8HV31 renealmia s |
| 26 | 121.5 | 3.9 | 392 | 2 Q63121 | Q63121 burkholderi |
| 27 | 121 | 3.8 | 520 | 2 Q672W9 | Q672W9 curculigo c |
| 28 | 120.5 | 3.8 | 479 | 2 Q75099 | Q75099 homo sapien |
| 29 | 119.5 | 3.8 | 390 | 2 Q7LCN3 | Q7LCN3 homo sapien |
| 30 | 119.5 | 3.8 | 411 | 2 Q9GZS9 | Q9GZS9 homo sapien |
| 31 | 119.5 | 3.8 | 518 | 2 Q9GHE3 | Q9GHE3 drymophila |

| | | | | | |
|----|-------|-----|-----|----------|---------------------|
| 32 | 119 | 3.8 | 257 | 2 Q794I5 | Q794I5 mus musculus |
| 33 | 119 | 3.8 | 472 | 2 Q88199 | Q88199 mus musculus |
| 34 | 118.5 | 3.8 | 358 | 2 Q7NM85 | Q7NM85 chromobacte |
| 35 | 118 | 3.7 | 395 | 2 Q9QUP4 | Q9QUP4 mus musculus |
| 36 | 118 | 3.7 | 520 | 2 Q6LA24 | Q6LA24 curculigo c |
| 37 | 117 | 3.7 | 515 | 2 Q8HV32 | Q8HV32 renealmia c |
| 38 | 117 | 3.7 | 559 | 2 Q9FGV3 | Q9FGV3 arabidopsis |
| 39 | 116.5 | 3.7 | 120 | 2 Q9DDK5 | Q9DDK5 mus musculus |
| 40 | 116.5 | 3.7 | 363 | 2 Q9VMC4 | Q9VMC4 drosophila |
| 41 | 116 | 3.7 | 363 | 2 Q9VMC4 | Q9VMC4 drosophila |
| 42 | 116 | 3.7 | 515 | 2 Q8HV29 | Q8HV29 renealmia t |
| 43 | 115 | 3.7 | 515 | 2 Q8HV34 | Q8HV34 renealmia a |
| 44 | 114 | 3.6 | 515 | 2 Q8HV65 | Q8HV65 etlingera e |
| 45 | 114 | 3.6 | 515 | 2 Q8HV21 | Q8HV21 siliquomom |
| | | | 515 | 2 Q8HV30 | Q8HV30 renealmia s |

ALIGNMENTS

RESULT 1
Q81ZU8 PRELIMINARY; PRT; 1222 AA.
AC Q81ZU8;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE NCAG1.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Gossens D., Del-Favero J., Van Broeckhoven C.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF480435; AA32895.1; .
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer 1; 1.
SQ SEQUENCE 1222 AA; 140272 MW; 156BE397B4E99AE9 CRC64;

Query Match 99.8%; Score 3142; DB 2; Length 1222;

Best Local Similarity 99.8%; Pred. No. 6.1e-243;

Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

| | | | |
|----|-----|---|------|
| Qy | 1 | MNRNGAMMDVDAHYKMFWDHGHGNSPMASIQEAOAEFKKRWTFVNVTFQWESIT | 60 |
| Db | 627 | MNRNGAMMDVDAHYKMFWDHGHGNSPMASIQEAOAEFKKRWTFVNVTFQWESIT | 686 |
| Qy | 61 | RIATVYFGPYINVSRCRFDSSNPGLOISLVNNTNTEHVSVITVDYHNLKTRFNYLGGFG | 120 |
| Db | 687 | RIATVYFGPYINVSRCRFDSSNPGLOISLVNNTNTEHVSVITVDYHNLKTRFNYLGGFG | 746 |
| Qy | 121 | ASVADQGITRFGIGTQAIKVPVHRDRIIPFGPKFNIAGVGLIICISIVILTFQWRPFLS | 180 |
| Db | 747 | ASVADQGITRFGIGTQAIKVPVHRDRIIPFGPKFNIAGVGLIICISIVILTFQWRPFLS | 806 |
| Qy | 181 | FRKLMRWILIVIALWFIELDVWSTCSQPICAKWTREAGSKSLSEGHMDLPDV | 240 |
| Db | 807 | FRKLMRWILIVIALWFIELDVWSTCSQPICAKWTREAGSKSLSEGHMDLPDV | 866 |
| Qy | 241 | ITSLPGSGAEILKOLFENSSDFLYIRVPTAYIDIPETELEIDSFVDAECWKVDIRSGHF | 300 |
| Db | 867 | ITSLPGSGAEILKOLFENSSDFLYIRVPTAYIDIPETELEIDSFVDAECWKVDIRSGHF | 926 |
| Qy | 301 | RLLRGLWQLSVQDTKHLQNIHLHPEHNRGLAQYFAMNKDKRKFRRRESLPEORSQWK | 360 |
| Db | 927 | RLLRGLWQLSVQDTKHLQNIHLHPEHNRGLAQYFAMNKDKRKFRRRESLPEORSQWK | 986 |
| Qy | 361 | AFDRDAEYIRALRRHLVYPSARVLSLSSGSWTLKHLFFQEVILGASMRALYIVRDPRW | 420 |
| Db | 987 | AFDRDAEYIRALRRHLVYPSARVLSLSSGSWTLKHLFFQEVILGASMRALYIVRDPRW | 1046 |

```
QY 421 IYSLMYSKPSLSXLSKNVPEHLAKLFKIEGKGKCNLNSGYAFEPYBPLRKELSKSNAV 480
|||||
Db 1047 IYSLMYSKPSLSXLSKNVPEHLAKLFKIEGKGKCNLNSGYAFEPYBPLRKELSKSNAV 1106
|||||
QY 481 SLLSHLWLANTAALRINTDOLLPTSQVLKPFEDIVHPFQKTTTRIFAFGIPLSPASLNQ 540
|||||
Db 1107 SLLSHLWLANTAALRINTDOLLPTSQVLKPFEDIVHPFQKTTTRIFAFGIPLSPASLNQ 1166
|||||
QY 541 ILFATSTNLFVLPYGEISPTNTNVKQNLPRDEIKLIENICWTLMDRLGYPKEMD 596
|||||
Db 1167 ILFATSTNLFVLPYGEISPTNTNVKQNLPRDEIKLIENICWTLMDRLGYPKEMD 1222
|||||
RESULT 2
Q9HAK5 PRELIMINARY; PRT; 755 AA.
ID Q9HAK5
AC Q9HAK5;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ11477.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Whole embryo;
RX PubMed=14702039; DOI=10.1038/ng1285;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahari K.,
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA Sudo H., Hosiiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,
RA Ninomiya K., Iehibashi T., Yamashita H., Murakawa K., Fujimori K.,
RA Tanai H., Kimata M., Watanabe M., Hiraoa S., Chiba Y., Ishida S.,
RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,
RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,
RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,
RA Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
RA Yoshihawa K., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
RA Moriya S., Nomiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe K., Kumagai A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
RT "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs.";
RL Nat. Genet. 36:40-45(2004).
RL EMBL; AK021539; BAB13840.1; -.
SQ SEQUENCE 755 AA; 86162 MW; A8104C69FA4D6250 CRC64;
Query Match 57.9%; Score 1821; DB 2; Length 755;
Best Local Similarity 99.4%; Pred. No. 3.7e-137;
Matches 342; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 MNRYNGAMVDVMDAHYKMFWDHGHGNSPMASIOEAQAEAFKRWTFQVNVTFQMESTIT 60
|||||
Db 411 MNRYNGAMVDVMDAHYKMFWDHGHGNSPMASIOEAQAEAFKRWTFQVNVTFQMESTIT 470
|||||
QY 61 RIAYVFYGPYINVSSCRFIDSSNPGLIQISLVNNTHTHVSVIVTDYHNLKTRFNLYLFGGFG 120
|||||
```

```
Db 471 RIAYVFYGPYINVSSCRFIDSSNPGLIQISLVNNTHTHVSVIVTDYHNLKTRFNLYLFGGFG 530
|||||
QY 121 ASVADQGOITRGLGTQALVKPVRHDIRIIFPPGKFNIAVGLICISLVILTFQWRFYLS 180
|||||
Db 531 ASVADQGOITRGLGTQALVKPVRHDIRIIFPPGKFNIAVGLICISLVILTFQWRFYLS 590
|||||
QY 181 FRKLNRWILLVIALWFIELDVWSTCSQOPICAKWTRTEAEGSKLSSEGHMDLPDVV 240
|||||
Db 591 FRKLNRWILLVIALWFIELDVWSTCSQOPICAKWTRTEAEGSKLSSEGHMDLPDVV 650
|||||
QY 241 ITSLSFGSAEILKQLFFNSSDFLYIRVPTAYIDIPETELEIDSFVDACEWKVSDIRSHP 300
|||||
Db 651 ITSLSFGSAEILKQLFFNSSDFLYIRVPTAYIDIPETELEIDSFVDACEWKVSDIRSHP 710
|||||
QY 301 RLLRGWLSQVDTKHLQNLHLHPENRGKLAQYFAMNKKRKK 344
|||||
Db 711 RLLRGWLSQVDTKHLQNLHLHPENRGKLAQYFAMNKKRKK 754
|||||
RESULT 3
Q9H951 PRELIMINARY; PRT; 480 AA.
ID Q9H951
AC Q9H951;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ13005.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=14702039; DOI=10.1038/ng1285;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahari K.,
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA Sudo H., Hosiiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,
RA Ninomiya K., Iehibashi T., Yamashita H., Murakawa K., Fujimori K.,
RA Tanai H., Kimata M., Watanabe M., Hiraoa S., Chiba Y., Ishida S.,
RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,
RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,
RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,
RA Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
RA Yoshihawa K., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
RA Moriya S., Nomiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe K., Kumagai A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
RT "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs.";
RL Nat. Genet. 36:40-45(2004).
RL EMBL; AK023067; BAB14387.1; -.
SQ SEQUENCE 480 AA; 54559 MW; AEP246851B83E9DB CRC64;
Query Match 29.2%; Score 920; DB 2; Length 480;
Best Local Similarity 100.0%; Pred. No. 4.4e-65;
Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MNRYNGAMVDVMDAHYKMFWDHGHGNSPMASIOEAQAEAFKRWTFQVNVTFQMESTIT 60
|||||
```

```

Db 308 MNRNGAMMDVDAHYKMFDFHGHGNSPMASIQEAOAEFKKRWTFQVNVVTFQWESTIT 367
Qy 61 RIAYVFGPYINVSSCFIDSSNPGLOISLVNNTNTEHVSVITDYHNLKTRFNVLGFGGF 120
Db 368 RIAYVFGPYINVSSCFIDSSNPGLOISLVNNTNTEHVSVITDYHNLKTRFNVLGFGGF 427
Qy 121 ASVADQOQITRFGGLGTOAIVKPVHRDRIIFPPGFKFNIAVGLILCISVILTF 173
Db 428 ASVADQOQITRFGGLGTOAIVKPVHRDRIIFPPGFKFNIAVGLILCISVILTF 480

RESULT 4
Q6P5Z3 PRELIMINARY; PRT; 784 AA.
ID Q6P5Z3
AC Q6P5Z3
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE C18orf4 protein (Fragment).
GN Name=C18orf4;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast;
RX MEDLINE=2388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Wax S.I., Wang J., Heish F.,
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Heiton E., Kettner M.C., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grinwood J., Schmutz J., Myers R.M., Burtfield Y.S.,
RA Krzyzanski M.I., Skalska U., Smalus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RL and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast;
RA Strauberg R.;
RL Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC062557; AAH62557.1; -.
FT NON TER 784
SQ SEQUENCE 784 AA; 89666 MW; 2A86FFC4DDCB846C CRC64;

Query Match 27.4%; Score 863; DB 2; Length 784;
Best Local Similarity 99.4%; Pred. No. 3.3e-60;
Matches 161; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MNRNGAMMDVDAHYKMFDFHGHGNSPMASIQEAOAEFKKRWTFQVNVVTFQWESTIT 60
Db 617 MNRNGAMMDVDAHYKMFDFHGHGNSPMASIQEAOAEFKKRWTFQVNVVTFQWESTIT 676
Qy 61 RIAYVFGPYINVSSCFIDSSNPGLOISLVNNTNTEHVSVITDYHNLKTRFNVLGFGGF 120
Db 677 RIAYVFGPYINVSSCFIDSSNPGLOISLVNNTNTEHVSVITDYHNLKTRFNVLGFGGF 736
Qy 121 ASVADQOQITRFGGLGTOAIVKPVHRDRIIFPPGFKFNIAVGL 162
Db 737 ASVADQOQITRFGGLGTOAIVKPVHRDRIIFPPGFKFNIAVGL 778

```

RESULT 5

```

Q9HAD7 PRELIMINARY; PRT; 314 AA.
ID Q9HAD7
AC Q9HAD7
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE Hypothetical protein FLJ11787.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Whole embryo;
RX PubMed=14702039; DOI=10.1038/ng1285;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahari K.,
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA Sudo H., Hosioki T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,
RA Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,
RA Tanai H., Kimata M., Watanabe M., Hiraoa K., Chiba Y., Ishida S.,
RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,
RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,
RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,
RA Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
RA Moriya S., Momiya H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Ohtani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togashi T., Oyama H., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
RT "Complete sequencing and characterization of 21,243 full-length human
RN cDNAs."
RL Nat. Genet. 36:40-45(2004).
RL EMBL; AK021849; BAB13912.1; -.
FT NON TER 314
SQ SEQUENCE 314 AA; 35955 MW; 722B5555DC065538 CRC64;

Query Match 19.5%; Score 615; DB 2; Length 314;
Best Local Similarity 100.0%; Pred. No. 7.6e-41;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 224 KXSLSSGHHMDLPDVVITSLPGSGAELKQLFNSDFLYRVTAVIDIPETELEIDS 283
Db 198 KXSLSSGHHMDLPDVVITSLPGSGAELKQLFNSDFLYRVTAVIDIPETELEIDS 257
Qy 284 FVDACEWKSVDIRSGHFRLLRGLQSLVQDTKLQHLHLEPNRGLAQYFAMNKD 340
Db 258 FVDACEWKSVDIRSGHFRLLRGLQSLVQDTKLQHLHLEPNRGLAQYFAMNKD 314

RESULT 6
Q9EQCO PRELIMINARY; PRT; 411 AA.
ID Q9EQCO
AC Q9EQCO
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)

```


Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.

RN (6)
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=2238257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Tohiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bobak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettaman M., Madan A., Rodriguez S., Sanchez A.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalka U., Smallus D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RA "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN (7)
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RA Strausberg R.;
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 RN (8)
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RA Strausberg R.;
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U65637; AAC28776.1; -;
 DR EMBL; AB003791; BAA24840.1; -;
 DR EMBL; AF090137; AAD19878.1; -;
 DR EMBL; AY339617; AAP88041.1; -;
 DR EMBL; BC022567; AAH22567.1; -;
 DR EMBL; BC028235; AAH28235.1; -;
 DR GO; GO:0016021; C: integral to membrane; TAS.
 DR GO; GO:0008146; P: sulfotransferase activity; TAS.
 DR GO; GO:0005976; P: polysaccharide metabolism; TAS.
 DR InterPro; IPR000863; Sulfotransferase.
 DR Pfam; PF00685; Sulfotransfer_1; 1.
 KW Transference.
 SQ SEQUENCE 411 AA; 46714 MW; F07DOA23B6338A09 CRC64;

Query Match 4.7%; Score 147; DB 2; Length 411;
 Best Local Similarity 18.2%; Pred. No. 0.0042;
 Matches 89; Conservative 58; Mismatches 149; Indels 194; Gaps 16;

Qy 187 WILVIALWFIEL-----LDVNSTC-----SOPICAKWTRTEAGSKSLSS 229
 Db 5 WKAVLLALASIAIQYTAIRFTAKSFHTCPLAENGLAERLC-----EESPTAYNLNR 59
 Qy 230 EGHMDLPDVVITSLPGSAILKQLFNSDFYIRVPTAYIDIPETELSDSFVACE 289
 Db 60 KTH-----ILILATTRSGSSFGVQLFQHLNDFYLFPEL----- 93
 Qy 290 WKVSDIRSGHFLRLGWLQSLVDQTKLHLNHLHLPNKGKLAQYFANMKDKKFKERE 349
 Db 94 -----YHVQNTLIPRFTQG-----KSPADRRV 115
 Qy 350 SLPEQRSMQKAFORDAEYI-----RALRHLVYPSARPV-----LS 387
 Db 116 MLCASRDLLRSYDCDLFLFENYKPPVNHHTDIFRGSARVLCSPVCPDPCPADLV 175
 Qy 388 LSSGSWTLLKHLFFQVFLGA-----SNRALLYIVRD 416
 Db 176 LEEGDCVRKCGLLNLTVAAACRERSHVAIKTVRVPVNDLRALVEDPRLNLKVIQLVRD 235

Qy 417 PRWIVSYMLYNKPSLYSLKQVPEHLAKLPKISGGKQKCNLNSGYAFEPYELPKELSKSK 476
 Db 236 PRGLASRSETFRDTY-----RLRWLYGTGR-----KPYNLVDVTLQ-L-TTVCEDF 279
 Qy 477 SNAVS--LLSHLWLANTAALRINTDLLPTSYOLVKFEEDIVHPQKTKTERIFAFGLPL- 533
 Db 280 SNSVSTGLMRPPWLKG-----KYMVLRYVEDLARNPMKTEELIYGLFGLPDL 325
 Qy 534 -----SPASLNQILFATSTNLFYLPYEGEISPTNTNVWKNQNLPRDEIKLIENI 581
 Db 326 SHVARMIONTRGDPTLGKHKGTVRN-----SAATAEKWRFLSYDIVAFAGNA 375
 Qy 582 CWTLMDRGLGY 591
 Db 376 CQOVLAQLGY 385

RESULT 8

Q9UL01 PRELIMINARY; PRT; 958 AA.
 AC Q9UL01;
 DT 01-MAY-2000 (T-EMBLrel. 13, Created)
 DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
 DT 05-JUL-2004 (T-EMBLrel. 27, Last annotation update)
 DE Squamous cell carcinoma antigen recognized by T cell (SART2 protein).
 DE protein.
 GN Name=SART-2;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 ON NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20143850; PubMed=10679095;
 RA Nakao M., Shichijo S., Imaizumi T., Inoue Y., Matsunaga K., Yamada A.,
 RA Kikuchi M., Tsuda N., Ohta K., Takamori S., Yamana H., Fujita H.,
 RA Itoh K.;
 RT "Identification of a gene coding for a new squamous cell carcinoma
 RL antigen recognized by the CTL";
 RL J. Immunol. 164:2565-2574 (2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RX MEDLINE=2238257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Brownstein M.J., Ustin T.B., Tohiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bobak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettaman M., Madan A., Rodriguez S., Sanchez A.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalka U., Smallus D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RA "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RA Strausberg R.;
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF098066; AAF00087.1; -;
 DR EMBL; BC039245; AAH39245.1; -;
 DR Genew; HGNC:21144; SART2.
 DR GO; GO:0005783; C: endoplasmic reticulum; TAS.


```

RESULT 10
Q794G9 PRELIMINARY; PRT; 483 AA.
AC Q794G9, 2004 (TrEMBLrel. 27, Created)
DT 05-JUN-2004 (TrEMBLrel. 27, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE N-acetylglucosamine-6-O-sulfotransferase (N-Acetylglucosamine 6-O-
DE Sulfotransferase-1).
GN Name=Gs6st-1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=whole embryos;
RX MEDLINE=98380482; PubMed=9712885; DOI=10.1074/jbc.273.35.22577;
RA Uchimura K., Muramatsu H., Kadomatsu K., Fan Q., Kurosawa N.,
RA Mitsuoka C., Kannagi R., Habuchi O., Muramatsu T.;
RT "Molecular cloning and characterization of an N-acetylglucosamine-6-O-
RT sulfotransferase.";
RL J. Biol. Chem. 273:22577-22583 (1998).
DR EMBL; AB011452; BAA32138.1; -
DR MGD; MGI:1891160; Chst2.
DR GO; GO:0005615; C:extracellular space; TAS.
DR GO; GO:0008146; F:sulfotransferase activity; IDA.
DR GO; GO:0006044; P:N-acetylglucosamine metabolism; TAS.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
DR Name=Chst2;
DR Transferrase.
DR KW Transferrase.
SQ SEQUENCE 483 AA; 52830 MW; 831FA08F5FEFD70E CRC64;

Query Match 4.3%; Score 134; DB 2; Length 483;
Best Local Similarity 26.2%; Pred. No. 0.059;
Matches 53; Conservative 36; Mismatches 81; Indels 32; Gaps 9;

QY 408 MRALYVRDPRAWIYMLYNKPSLY--SLKNVPE-----HLAKLFKIEGK-GKCNLNS 459
DB 278 LKVIHLVRDPRA-VASSRIRSRHGLIRESLQVRSRDPRAHRMPPLEAAGHKLGAKKEGM 336
QY 460 GYAFEYELPKELSKSNVSLSHLWLAALRINTDLPSTYQVLFKEDIVHPQ 519
DB 337 GGPADYHAL-----GAMEVICNSMAKTLQALQ-PPDWLQGHYLVRYEDLVGDPV 386
QY 520 KTERIFAFGLPIPLSPA----SLNQILPATSTNLFYLYPEGEISPTN----TNVWKQNL 571
DB 387 KTLRRVYDFVGLLVSPENEPQALNWTSGSSGSKPPFV-----VSARNATQANAWRTALT 441
QY 572 RDEIKLIENICWTLMRLGYPK 593
DB 442 FQIQKVEEFCYQPMVLYGER 463

RESULT 11
O88276 PRELIMINARY; PRT; 530 AA.
AC O88276, 1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DE N-acetylglucosamine-6-O-sulfotransferase long form.
DE N-acetylglucosamine-6-O-sulfotransferase long form.
GN Name=Chst2;
OS Mus musculus (Mouse).

```

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=whole embryos;
RX MEDLINE=98380482; PubMed=9712885; DOI=10.1074/jbc.273.35.22577;
RA Uchimura K., Muramatsu H., Kadomatsu K., Fan Q., Kurosawa N.,
RA Mitsuoka C., Kannagi R., Habuchi O., Muramatsu T.;
RT "Molecular cloning and characterization of an N-acetylglucosamine-6-O-
RT sulfotransferase.";
RL J. Biol. Chem. 273:22577-22583 (1998).
DR EMBL; AB011452; BAA32138.1; -
DR MGD; MGI:1891160; Chst2.
DR GO; GO:0005615; C:extracellular space; TAS.
DR GO; GO:0008146; F:sulfotransferase activity; IDA.
DR GO; GO:0006044; P:N-acetylglucosamine metabolism; TAS.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
DR Name=Chst2;
DR Transferrase.
DR KW Transferrase.
SQ SEQUENCE 530 AA; 57814 MW; A113E1B735C363EC CRC64;

Query Match 4.3%; Score 134; DB 2; Length 530;
Best Local Similarity 26.2%; Pred. No. 0.067;
Matches 53; Conservative 36; Mismatches 81; Indels 32; Gaps 9;

QY 408 MRALYVRDPRAWIYMLYNKPSLY--SLKNVPE-----HLAKLFKIEGK-GKCNLNS 459
DB 325 LKVIHLVRDPRA-VASSRIRSRHGLIRESLQVRSRDPRAHRMPPLEAAGHKLGAKKEGM 383
QY 460 GYAFEYELPKELSKSNVSLSHLWLAALRINTDLPSTYQVLFKEDIVHPQ 519
DB 384 GGPADYHAL-----GAMEVICNSMAKTLQALQ-PPDWLQGHYLVRYEDLVGDPV 433
QY 520 KTERIFAFGLPIPLSPA----SLNQILPATSTNLFYLYPEGEISPTN----TNVWKQNL 571
DB 434 KTLRRVYDFVGLLVSPENEPQALNWTSGSSGSKPPFV-----VSARNATQANAWRTALT 488
QY 572 RDEIKLIENICWTLMRLGYPK 593
DB 489 FQIQKVEEFCYQPMVLYGER 510

RESULT 12
O80WV3 PRELIMINARY; PRT; 530 AA.
AC O80WV3, 2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Chst2 protein.
GN Name=Chst2;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Berge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B.,
RA Brownstein M.J., Ussin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Hulyk S.W.,
RA Fahey J., Helton E., Kettaman M., Madan A., Rodríguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

```

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalek U., Smallos D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strausberg R.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC051963; AAH51963.1; -.
DR MGD; MGI:1891160; Chst2.
DR GO; GO:0005615; C:extracellular space; TAS.
DR GO; GO:0008146; F:sulfotransferase activity; IDA.
DR GO; GO:0006044; P:N-acetylglucosamine metabolism; TAS.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
SQ SEQUENCE 530 AA; 57828 MW; 275363BF15440730 CRC64;

Query Match 4.3%; Score 134; DB 2; Length 530;
Best Local Similarity 26.2%; Pred. No. 0.067;
Matches 53; Conservative 36; Mismatches 81; Indels 32; Gaps 9;

QY 408 MRALYIVRDPRAWIYSLYNSKPSLY--SLKNVPE-----HLAKLFKIEGK-GKCNLNS 459
DB 325 LKVIHLVRDPA-VASSRIRSRHGLIRESLQVRSRDPRAHRMPFLEAGHKLGAKEGM 383
QY 460 GYAFEPYELRKLKSKSNVSLSHLWLANTAAALRINTDLLPTSQYLVKPEFEDIVHFPQ 519
DB 384 GGPADYHAL-----GAMEVICNSMAKTLQTLQ-PPDWLQGHYLVVRYEDLVGDPV 433
QY 520 KTERIFAFIPIGLPSA-----SLNQILFATSNLFLPYLVEGEISPTN-----TNVWKQNL 571
DB 434 KTLRRVYDFVGLLVSPMEQFALNMTSGSGSSKPFV-----VSARNATQAAANWRTALT 488
QY 572 RDEIKLIENICWTLMRLGYPK 593
DB 489 FQIKQVEEFCYQPMVILGYER 510

RESULT 13
QYUED5 ID QYUED5 PRELIMINARY; PRT; 483 AA.
AC QYUED5; 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE N-acetylglucosamine-6-O-sulfotransferase (GlcNAc6ST).
GN Name=GN6ST; Synonyms=CHST2;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98391845; PubMed=9722682;
RA Uchimura K., Muramatsu H., Kaname T., Ogawa H., Yamakawa T., Fan Q.,
RA Mitsuoka C., Kannagi R., Habuchi O., Yokoyama I., Yamamura K.,
RA Ozaki T., Nakagawara A., Kadamatsu K., Muramatsu T.;
RT "Human N-acetylglucosamine-6-O-sulfotransferase involved in the
RT biosynthesis of 6-sulfo sialyl Lewis X: molecular cloning, chromosomal
RT mapping, and expression in various organs and tumor cells.";
RL J. Biochem. 124:670-678(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=20498786; PubMed=11042394; DOI=10.1016/S0304-4165(00)00136-7;
RA Sakaguchi H., Kitagawa H., Sugahara K.;
RT "Functional expression and genomic structure of human N-
RT acetylglucosamine-6-O-sulfotransferase that transfers sulfate to b-N-

RT acetylglucosamine at the nonreducing end of an N-acetylglucosamine
RT sequence";
RL Biochim. Biophys. Acta 1523:269-276(2000).
DR EMBL; AB014679; BAA34265.2; -.
DR EMBL; AB021124; BAB16886.1; -.
DR EMBL; AB021125; BAB16887.1; -.
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
KW Transferase.
SQ SEQUENCE 483 AA; 52787 MW; E27797D444931BA18 CRC64;

Query Match 4.2%; Score 133; DB 2; Length 483;
Best Local Similarity 26.2%; Pred. No. 0.071;
Matches 53; Conservative 36; Mismatches 81; Indels 32; Gaps 9;

QY 408 MRALYIVRDPRAWIYSLYNSKPSLY--SLKNVPE-----HLAKLFKIEGK-GKCNLNS 459
DB 278 LKVIHLVRDPA-VASSRIRSRHGLIRESLQVRSRDPRAHRMPFLEAGHKLGAKEGV 336
QY 460 GYAFEPYELRKLKSKSNVSLSHLWLANTAAALRINTDLLPTSQYLVKPEFEDIVHFPQ 519
DB 337 GGPADYHAL-----GAMEVICNSMAKTLQTLQ-PPDWLQGHYLVVRYEDLVGDPV 386
QY 520 KTERIFAFIPIGLPSA-----SLNQILFATSNLFLPYLVEGEISPTN-----TNVWKQNL 571
DB 387 KTLRRVYDFVGLLVSPMEQFALNMTSGSGSSKPFV-----VSARNATQAAANWRTALT 441
QY 572 RDEIKLIENICWTLMRLGYPK 593
DB 442 FQIKQVEEFCYQPMVILGYER 463

RESULT 14
QYV4C5 ID QYV4C5 PRELIMINARY; PRT; 530 AA.
AC QYV4C5; 09GZNS; 09Y6P2;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Long form of N-acetylglucosamine-6-O-sulfotransferase (GlcNAc6ST)
DE (Carbohydrate sulfotransferase 2).
GN Name=GN6ST; Synonyms=CHST2;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98391845; PubMed=9722682;
RA Uchimura K., Muramatsu H., Kaname T., Ogawa H., Yamakawa T., Fan Q.,
RA Mitsuoka C., Kannagi R., Habuchi O., Yokoyama I., Yamamura K.,
RA Ozaki T., Nakagawara A., Kadamatsu K., Muramatsu T.;
RT "Human N-acetylglucosamine-6-O-sulfotransferase involved in the
RT biosynthesis of 6-sulfo sialyl Lewis X: molecular cloning, chromosomal
RT mapping, and expression in various organs and tumor cells.";
RL J. Biochem. 124:670-678(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Umbilical vein endothelium;
RX MEDLINE=99168906; PubMed=10049591; DOI=10.1006/geno.1998.5653;
RA Li X., Tedder T.F.;
RT "CHST1 and CHST2 sulfotransferases expressed by human vascular
RT endothelial cells: cDNA cloning, expression, and chromosomal
RT localization";
RL Genomics 55:345-347(1999).
DR EMBL; AB014680; BAA34266.2; -.
DR EMBL; AF083066; AAD20981.1; -.
DR Genew; HGNC:1970; CHST2.
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.

DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
KW Transferase.
SQ SEQUENCE 530 AA; 57857 MW; A82CA227B9D5651B CRC64;

Query Match 4.2%; Score 133; DB 2; Length 530;
Best Local Similarity 26.2%; Pred. No. 0.077;
Matches 53; Conservative 36; Mismatches 81; Indels 32; Gaps 9;

QY 408 MRALYIVRDPRAWIYMLYNSKPSLY--SLKNVPE-----HLAKLFKIEGCK-GKCNLNS 459
DB 325 LKVHLVRDPA--VASSRISSRHGLIESQVVSRRDPRAHRMPFLAAGHKLGAKEGV 383
QY 460 GYAFEPYELRKLKSKSNVSLSHLWLAALRINTDLLPTSQYLVKFEDIVHFPQ 519
DB 384 GGPADYHAL-----GAMEVICNSMAKLTQALQ-PPDWLQGHYLVVRYEDLVGDPV 433
QY 520 KTERIFAFGLIPUSPA-----SLNQILFATSTNLFYLPYEGEISPTN-----TNWKNQNL 571
DB 434 KTLRRVYDFVGLLVSPMEQFALNMTSGSGSSKPFV-----VSARNATQAANAWRTALT 488
QY 572 RDEIKLIENICWTLMDRIGYPK 593
DB 489 FQIKQVEEFCYQPMVILGYER 510

RESULT 15
Q677T4 PRELIMINARY; PRT; 304 AA.

AC Q677T4
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Hypothetical protein.
OS Lymphocystis disease virus - isolate China.
OC Viruses; dsDNA viruses, no RNA stage; Iridoviridae; Lymphocystivirus.
OX NCBI_TaxID=256729;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=15194775; DOI=10.1128/JVI.78.13.6982-6994.2004;
RA Zhang Q.X., Xiao J., Xie J., Li Z.Q., Gui J.F.;
RT "Complete Genome Sequence of Lymphocystis Disease Virus Isolated from China."
RL J. Virol. 78:6982-6994(2004).
DR EMBL; AY380826; AAU11023.1; -
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
KW Hypothetical protein.
SQ SEQUENCE 304 AA; 35898 MW; 63C9BDBFB2145C22 CRC64;

Query Match 4.1%; Score 129; DB 2; Length 304;
Best Local Similarity 21.1%; Pred. No. 0.077;
Matches 86; Conservative 62; Mismatches 126; Indels 134; Gaps 22;

QY 188 ILILVIALMFTLELDVWSTCSQPCAKWTRTEARGSKSLSEGHMDLPDWITSLPGS 247
DB 8 MLLIILLWL-----IW-IFQRPVI-----YEKPNK-----ILLITTRS 43
QY 248 GAETLKQLFFNSSDFLYIRVPTAYIDIPETELEIDSFVDACEWKVSDIRSGHFRLLRGWL 307
DB 44 GSSFLGEIFNRSRSDVFLFEPL-----WHLNEIYEVYKVLKALF 83
QY 308 Q-SLVQDTKLHLQNH-----LHEPNRGLAQYFAMNKKKKKFKRESLP-----EQR 355
DB 84 NCELVLRLKYLTONFFFKRNYSKALCKP--GKTCIY-GLNGDRKYTCGLKCPQLNDIA 140
QY 356 SOMKGFORDAEYIRALRHLVYPSARPVLSLSSGSMWTLKLFHFQEVLGASMEALYIVR 415
DB 141 SLYCQTF--DTWIKTVR-----IRNKQALEL-----WTQFDIKIHLVR 179
QY 416 DPRAWIYMLYNSKPSLYSLKNVPEHLAKLFKIEGCKGNLNSGYAFEPYELRKLKLS 475
DB 180 DPRG-----SFSNK-----IKTF-----NRDYNF-----KQIAKI 204

QY 476 KSNVSLLSHLWLAALRINTDLLPTSQYLVKFEDIVHFPQKTTTERIPAFGLIPLS 535
DB 205 QDDIDIYETL-----KDRL--GYLLKYEDLIINPKELPHLFSFCELMFE 250
QY 536 ASLQNLFPATSTNLFYLPYE-GEISPTNTNVWKNLPRDEIKLIENIC 582
DB 251 EVLKTIDRLTNTNRSG-PYAIGKIKPSG---WKTQLSTFKIDMIENAC 294

Search completed: June 23, 2005, 08:49:00
Job time : 67.4049 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2005, 08:32:43 ; Search time 42.6885 Seconds
(without alignments)
3578.727 Million cell updates/sec

Title: US-10-697-828-13

Perfect score: 2056

Sequence: 1 MWLPRVSSTAVTALLAQTF.....LPRGLNGFTWASSTASHPRN 395

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : A_Geneseq_16Dec04:*

1: Geneseqp1980s:*

2: Geneseqp1990s:*

3: Geneseqp2000s:*

4: Geneseqp2001s:*

5: Geneseqp2002s:*

6: Geneseqp2003as:*

7: Geneseqp2003bs:*

8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|------------|--------------------|
| 1 | 2056 | 100.0 | 395 | 4 AAY72640 | Aay72640 Human gly |
| 2 | 2056 | 100.0 | 395 | 5 ABB81554 | Abb81554 Human cor |
| 3 | 2056 | 100.0 | 395 | 5 AAE15438 | Aae15438 Human dru |
| 4 | 2056 | 100.0 | 395 | 7 ADI21086 | Adi21086 Novel hum |
| 5 | 2056 | 100.0 | 395 | 8 ADL61235 | Adl61235 Human tyr |
| 6 | 1934 | 94.1 | 395 | 5 ABB81555 | Abb81555 Consensus |
| 7 | 1729.5 | 84.1 | 390 | 5 AAY72639 | Aay72639 Human gly |
| 8 | 1729.5 | 84.1 | 390 | 5 ABB81556 | Abb81556 Human int |
| 9 | 1708 | 83.1 | 395 | 4 AAY72638 | Aay72638 Mouse gly |
| 10 | 1708 | 83.1 | 395 | 5 AAU11275 | Aau11275 Murine in |
| 11 | 1694.5 | 82.4 | 418 | 5 ABB81557 | Abb81557 Mouse int |
| 12 | 1620 | 78.8 | 418 | 3 AAB41947 | Aab41947 Human ORF |
| 13 | 1538 | 74.8 | 394 | 7 ADJ70405 | Adj70405 Human hea |
| 14 | 1024 | 49.8 | 386 | 4 AAM93309 | Aam93309 Human pol |
| 15 | 1024 | 49.8 | 386 | 8 ADL30784 | Adl30784 Human pro |
| 16 | 1019 | 49.6 | 386 | 2 AAY39918 | Aay39918 Human gly |
| 17 | 1018 | 49.5 | 380 | 5 AAU11274 | Aau11274 Human L-s |
| 18 | 981 | 47.7 | 386 | 3 AAY79219 | Aay79219 Human tra |
| 19 | 977 | 47.5 | 388 | 2 AAY39919 | Aay39919 Mouse gly |
| 20 | 808 | 39.3 | 169 | 5 ABB81558 | Abb81558 Human cor |
| 21 | 755 | 36.7 | 169 | 5 ABB81559 | Abb81559 Human int |
| 22 | 663 | 32.2 | 483 | 2 AAY31656 | Aay31656 Mouse N-a |
| 23 | 657.5 | 32.0 | 530 | 4 AAB95367 | Aab95367 Human pro |
| 24 | 657.5 | 32.0 | 530 | 8 ADQ18590 | Adq18590 Human sof |
| 25 | 653.5 | 31.8 | 484 | 2 AAY31657 | Aay31657 Human N-a |

ALIGNMENTS

RESULT 1

AAAY72640

ID AAY72640 standard; protein; 395 AA.

XX AC AAY72640;

XX XX 02-MAY-2001 (first entry)

XX DE Human glycosyl sulfotransferase-4beta (GST-4beta).

XX KW Human; glycosyl sulfotransferase-4beta; GST-4beta; immunosuppressive; therapy; selectin binding inhibitor; gene therapy; inflammation; systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes; polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis; glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenalitis; Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia; demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis; myocarditis; adult respiratory distress syndrome; eczema; psoriasis; asthma; hypersensitivity; rheumatic fever; tissue rejection; chromosome 16q23.1.

XX OS Homo sapiens.

XX PN WO200106015-A1.

XX PD 25-JAN-2001.

XX PF 19-JUL-2000; 2000WO-US019741.

XX PR 20-JUL-1999; 99US-0144694P.

XX PR 13-JUN-2000; 2000US-00593828.

XX PA (REGC) UNIV CALIFORNIA.

XX PI Rosen SD, Lee JK, Hemmerich S;

XX PI WPI; 2001-138471/14.

XX PI N-PSDB; AAD02697, AAD02700.

XX PT New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for diagnostic and therapeutic agent screening applications.

XX PS Claim 3; Fig 4B; 128pp; English.

XX CC The present sequence is human glycosyl sulfotransferase-4beta (GST-4 beta). GST-4 gene is found on chromosome 16q23.1. GST is a type 2

XX CC membrane protein useful for inhibiting a binding event between a selectin and a selectin ligand, which comprises contacting the selectin with a non

26 653.5 31.8 531 5 AAU69414
27 653.5 31.8 531 8 ADI14283
28 653 31.8 128 7 ADI21571
29 614 29.9 486 6 ABP56121
30 578.5 28.1 479 7 ABM85237
31 577.5 28.1 499 6 ABR41139
32 576.5 28.0 479 2 AAW52863
33 569 27.7 481 7 ABM85236
34 555.5 27.0 411 2 AAW61100
35 555.5 27.0 411 5 AAE25356
36 555.5 27.0 411 6 ABU03503
37 555.5 27.0 411 7 ADJ68589
38 555.5 27.0 411 8 ADQ18725
39 554.5 27.0 410 7 ADN95522
40 546.5 26.6 504 8 ABM85022
41 532.5 25.9 458 2 AAW06480
42 521.5 25.4 171 5 ABB81560
43 377 18.3 169 5 ABB81561
44 364.5 17.7 174 5 ABB81563
45 348 16.9 179 5 ABB81562

Aau69414 Lung amal
Adri4283 Human NF-
Adi21571 Novel hum
Abp56121 Human cho
Abm85237 Human pro
Abr41139 Human DIT
Aaw52863 Glycosami
Abm85236 Mouse pro
Aaw61100 Keratan s
Aae25356 Human cho
Abu03503 Angiogene
Adj68589 Human hea
Adq18725 Human sof
Adn95522 Human BEC
Abm85022 Human dia
Aaw06480 Chick cho
Abb81560 Human hig
Abb81561 Human N-a
Abb81563 Human cho
Abb81562 Human ker

CC -sulphated selectin ligand, GST and a small molecular agent that inhibits
 CC the sulphation activity of GST. GST is also useful in inhibiting a
 CC selectin mediated binding event. GST is useful in gene therapy to treat
 CC disorders such as acute or chronic inflammation, systemic lupus
 CC erythematosus (SLE), rheumatoid arthritis, polyarteritis nodosa,
 CC polymyositis, dermatomyositis, systemic sclerosis, diabetes,
 CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, pernicious
 CC disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious
 CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis, distress
 CC dermatitis, myocarditis, regional enteritis, adult respiratory distress
 CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
 CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
 CC during transplantation
 XX
 SQ Sequence 395 AA;

Query Match 100.0%; Score 2056; DB 4; Length 395;
 Best Local Similarity 100.0%; Pred. No. 3.1e-207;
 Matches 395; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MWLPRVSTAVTALLLAQTLLFLVSRPGSPAGGEARVHVLVLSWRSGSFFVGQLF 60
 DB 1 MWLPRVSTAVTALLLAQTLLFLVSRPGSPAGGEARVHVLVLSWRSGSFFVGQLF 60
 QY 61 NOHPDVFYLMPEPAWHVMTTLLSQGSAATLHMAVRLVRSVFLCDMDVFDAYLPWRRNLSL 120
 DB 61 NOHPDVFYLMPEPAWHVMTTLLSQGSAATLHMAVRLVRSVFLCDMDVFDAYLPWRRNLSL 120
 QY 121 FQWAVSRALCSPACSAFPRGAISSSEAVCKPLCARQSFLLAREACRSYSHVVLKEVRFN 180
 DB 121 FQWAVSRALCSPACSAFPRGAISSSEAVCKPLCARQSFLLAREACRSYSHVVLKEVRFN 180
 QY 181 LOVLYPLSDPALNRIVHLVRDPAVLRSREQTAKALARDNGIVLGTNGTWVEADPGLR 240
 DB 181 LOVLYPLSDPALNRIVHLVRDPAVLRSREQTAKALARDNGIVLGTNGTWVEADPGLR 240
 QY 241 VREVCRRSHVRIAEAAATLKPPPLGRYRLVRFPEDLAREPLAEIRALYAFTGLSLTPQLE 300
 DB 241 VREVCRRSHVRIAEAAATLKPPPLGRYRLVRFPEDLAREPLAEIRALYAFTGLSLTPQLE 300
 QY 301 AWIHNITHGSGPGARREAFKTSRRNALNVSQAWRHALPFAKIRRVQELCAGALQLLGYRP 360
 DB 301 AWIHNITHGSGPGARREAFKTSRRNALNVSQAWRHALPFAKIRRVQELCAGALQLLGYRP 360
 QY 361 VYSEDEQRNALDLVLPRLGNGFTWASSTASHPRN 395
 DB 361 VYSEDEQRNALDLVLPRLGNGFTWASSTASHPRN 395

RESULT 2

ABB81554
 ID ABB81554 standard; protein; 395 AA.

XX ABB81554;

DT 05-SEP-2002 (first entry)

XX Human corneal N-acetylglucosamine-6-sulfotransferase SEQ ID NO:2.

XX Human; N-acetylglucosamine-6-sulfotransferase; enzyme; GlcNAc6ST;
 KW corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
 KW ophthalmological; chromosome 16q22.

XX Homo sapiens.

XX US2002061562-A1.

XX 23-MAY-2002.

XX 09-AUG-2001; 2001US-00927602.

XX 11-AUG-2000; 2000US-00638211.

PR 11-AUG-2000; 2000US-0325773P.

XX (FUKU/) FUKUDA M.N.
 PA (AKAM/) AKAMA T.O.
 XX Fukuda MN, Akama TO;
 XX WPI; 2002-507643/54.
 DR N-PSDB; ABB89506.
 XX New nucleic acid encoding corneal N-acetylglucosamine-6-sulfotransferase,
 PT useful for treatment, monitoring and diagnosis of macular corneal
 PT dystrophy.
 XX
 PS Claim 13; Fig 1A-D; 69pp; English.
 XX The present sequence represents human corneal N-acetylglucosamine-6-
 CC sulfotransferase (I), which is able to catalyze sulfation of keratan
 CC sulfate (KS). Also described is a method for monitoring the effect of
 CC treatments for macular corneal dystrophy (MCD), and detecting
 CC susceptibility to MCD. (I) is located to chromosome 16q22, and has
 CC ophthalmological activity. (I) can be used to treat or prevent macular
 CC corneal dystrophy types I or II. (I) makes possible treatment of MCD
 CC without requiring keratoplasty or keratectomy
 XX
 SQ Sequence 395 AA;
 Query Match 100.0%; Score 2056; DB 5; Length 395;
 Best Local Similarity 100.0%; Pred. No. 3.1e-207;
 Matches 395; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MWLPRVSTAVTALLLAQTLLFLVSRPGSPAGGEARVHVLVLSWRSGSFFVGQLF 60
 DB 1 MWLPRVSTAVTALLLAQTLLFLVSRPGSPAGGEARVHVLVLSWRSGSFFVGQLF 60
 QY 61 NOHPDVFYLMPEPAWHVMTTLLSQGSAATLHMAVRLVRSVFLCDMDVFDAYLPWRRNLSL 120
 DB 61 NOHPDVFYLMPEPAWHVMTTLLSQGSAATLHMAVRLVRSVFLCDMDVFDAYLPWRRNLSL 120
 QY 121 FQWAVSRALCSPACSAFPRGAISSSEAVCKPLCARQSFLLAREACRSYSHVVLKEVRFN 180
 DB 121 FQWAVSRALCSPACSAFPRGAISSSEAVCKPLCARQSFLLAREACRSYSHVVLKEVRFN 180
 QY 181 LOVLYPLSDPALNRIVHLVRDPAVLRSREQTAKALARDNGIVLGTNGTWVEADPGLR 240
 DB 181 LOVLYPLSDPALNRIVHLVRDPAVLRSREQTAKALARDNGIVLGTNGTWVEADPGLR 240
 QY 241 VREVCRRSHVRIAEAAATLKPPPLGRYRLVRFPEDLAREPLAEIRALYAFTGLSLTPQLE 300
 DB 241 VREVCRRSHVRIAEAAATLKPPPLGRYRLVRFPEDLAREPLAEIRALYAFTGLSLTPQLE 300
 QY 301 AWIHNITHGSGPGARREAFKTSRRNALNVSQAWRHALPFAKIRRVQELCAGALQLLGYRP 360
 DB 301 AWIHNITHGSGPGARREAFKTSRRNALNVSQAWRHALPFAKIRRVQELCAGALQLLGYRP 360
 QY 361 VYSEDEQRNALDLVLPRLGNGFTWASSTASHPRN 395
 DB 361 VYSEDEQRNALDLVLPRLGNGFTWASSTASHPRN 395
 RESULT 3
 AAE15438
 ID AAE15438 standard; protein; 395 AA.
 XX AAE15438;
 XX
 DT 12-MAR-2002 (first entry)
 XX Human drug metabolising enzyme (DME) -5.
 XX Human; drug metabolising enzyme; gene therapy; autoimmune disorder;
 KW inflammatory disorder; acquired immune deficiency syndrome; infection;
 KW AIDS; adult respiratory distress syndrome; cell proliferative disorder;
 KW allergy; anaemia; conjunctivitis; actinic keratosis; arteriosclerosis;

KW cancer; endocrine disorder; hypothalamus disorder; pituitary disorder;
 KW gastrointestinal disorder; metabolic disorder; developmental disorder;
 KW liver disorder; iritis; cystic fibrosis; Addison's disease; retinitis;
 KW goitre; diabetes; dysphagia; gastric carcinoma; anorexia; eye disorder;
 KW DME-5.
 XX Homo sapiens.

XX Key Location/Qualifiers
 FH Peptide 1..35
 FT /label= Signal_peptide
 FT Peptide 1..32
 FT /label= Signal_peptide
 FT Protein 33..395
 FT /note= "Human mature DME-5 protein"
 FT Protein 36..395
 FT /note= "Human mature DME-5 protein"

XX WO200179468-A2.

XX 25-OCT-2001.

XX 12-APR-2001; 2001WO-US011869.

XX 13-APR-2000; 2000US-0197590P.

PR 19-APR-2000; 2000US-0198403P.

PR 28-APR-2000; 2000US-0200185P.

PR 05-MAY-2000; 2000US-0202234P.

PR 11-MAY-2000; 2000US-0203509P.

XX (INCY-) INCYTE GENOMICS INC.

XX POLICKY JL, Hafalia A, Burford N, Ring HZ, Lal P, Tribouley CM;

PI Yao MG, Yue H, Tang YT, Patterson C, Das D, Sanjanwala MS;

PI Gandhi AR, Reddy R, Khan FA, Baughn MR, Ramkumar J, Griffin JA;

PI Au-Young J;

XX WPI; 2002-066363/09.

DR N-PSDB; AAD24670.

XX Novel isolated human drug metabolizing enzymes referred as DME 1-10,

PT useful for diagnosing, treating, or preventing disorders associated with

PT aberrant expression of DME such as allergy, anemia, asthma, infertility.

XX Claim 1a; Page 131-132; 143pp; English.

XX The invention relates to human drug metabolising enzymes referred as DME

CC and nucleic acid molecules encoding such enzymes. Polynucleotides of the

CC invention are useful for assessing toxicity of test compounds and in gene

CC therapy. Sequences of the invention are useful in the diagnosis,

CC prevention and treatment of autoimmune/inflammatory disorders such as

CC acquired immune deficiency syndrome (AIDS), adult respiratory distress

CC syndrome, allergies, anaemia, atherosclerosis, asthma, autoimmune

CC haemolytic anaemia, contact dermatitis, Crohn's disease,

CC Hashimoto's thyroiditis, irritable bowel syndrome, gout, Grave's disease,

CC osteoarthritis, osteoporosis, psoriasis, systemic lupus erythematosus,

CC rheumatoid arthritis, scleroderma, ulcerative colitis, vitellitis, viral,

CC bacterial, fungal, parasitic, protozoal, helminthic infections; cell

CC proliferative disorders such as actinic keratosis, arteriosclerosis,

CC atherosclerosis, Duchenne and Becker dystrophy, hepatitis, Cushing's

CC syndrome, cancers, myelodysplastic syndrome, epilepsy, endocrine

CC disorders such as disorders of the hypothalamus and pituitary resulting

CC from lesions such as primary brain tumours, adenomas, infarction

CC associated with pregnancy, aneurysms, vascular malformations; eye

CC disorders such as conjunctivitis, iritis, retinitis, glaucoma, pigmentosa

CC ; metabolic disorders such as Addison's disease, cystic fibrosis,

CC diabetes, goitre, glycogen storage diseases, hypercholesterolaemia,

CC hyperthyroidism, hypoglycaemia, lipid myopathies, Menkes syndrome,

CC mannosidosis, obesity; gastrointestinal disorders such as dysphagia,

CC gastric carcinoma, anorexia, nausea, gastroenteritis,

CC hyperbilirubinaemia, emesis, cirrhosis, diarrhoea, jaundice, Reye's

CC syndrome, peliosis hepatitis, hepatic vein thrombosis and developmental

XX

CC disorders. The present sequence is human DME-5 protein

XX

SQ Sequence 395 AA;

Query Match 100.0%; Score 2056; DB 5; Length 395;

Best Local Similarity 100.0%; Pred No. 3.1e-207; Indels 0; Gaps 0;

Matches 395; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MMLPRVSTAVTALLLAQTFLFLVSRPGSPSPAGGEARVHVLVLSWRSGSFVQGLF 60

DB 1 MMLPRVSTAVTALLLAQTFLFLVSRPGSPSPAGGEARVHVLVLSWRSGSFVQGLF 60

QY 61 NQHPDVFYLMPEAWHVTTLTSSQSAATLHMAVRLVRSVFLCDMDVDFDAYLPWRRNLSDL 120

DB 61 NQHPDVFYLMPEAWHVTTLTSSQSAATLHMAVRLVRSVFLCDMDVDFDAYLPWRRNLSDL 120

QY 121 FQWAVSRALCSPACSAFPRGATSSSAVCKPLCAROSFTLAREACRSYSHVVLKEVRFN 180

DB 121 FQWAVSRALCSPACSAFPRGATSSSAVCKPLCAROSFTLAREACRSYSHVVLKEVRFN 180

QY 181 LQVLYPLSDPALNLRIVHLVRDPRAVLRSEQTAKALARDNGIVLTGNTGWVADPGLR 240

DB 181 LQVLYPLSDPALNLRIVHLVRDPRAVLRSEQTAKALARDNGIVLTGNTGWVADPGLR 240

QY 241 VREVCRSVRIAEAAATLKPPPLRGYRLVRPDLAREPLAEIRALYATGLSLTPGLE 300

DB 241 VREVCRSVRIAEAAATLKPPPLRGYRLVRPDLAREPLAEIRALYATGLSLTPGLE 300

QY 301 AWIHNTHGSGPGARREAFKTSRNALNVSQAWRHALPFAKIRRVQELCAGALQLLGYRP 360

DB 301 AWIHNTHGSGPGARREAFKTSRNALNVSQAWRHALPFAKIRRVQELCAGALQLLGYRP 360

QY 361 VYSEDEQRNALDLVLPRLGNGFTWASSTASHPRN 395

DB 361 VYSEDEQRNALDLVLPRLGNGFTWASSTASHPRN 395

RESULT 4

ADI21086

ID ADI21086 standard; protein; 395 AA.

XX ADI21086;

XX ADI21086;

DT 15-APR-2004 (first entry)

XX Novel human protein #61.

XX forensic; nutritional source; damaged tissue; diseased tissue;

XX myeloid cell disorder; lymphoid cell disorder;

XX bone cartilage tissue growth; tendon tissue growth;

XX ligament tissue growth; nerve tissue growth; regeneration; wound healing;

XX tissue repair; tissue replacement; burn; incision; ulcer; cancer; human.

XX Homo sapiens.

XX WO2003025148-A2.

XX 27-MAR-2003.

XX 19-SEP-2002; 2002WO-US029964.

XX 19-SEP-2001; 2001US-0323739P.

XX 13-SEP-2002; 2002US-00323739.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;

XX Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang D;

XX Haley-Vicente D;

XX WPI; 2003-354603/33.

XX N-PSDB; ADI21802.

XX

PT New polynucleotides and secreted proteins, useful for treating myeloid or
PT lymphoid cell disorders, in bone cartilage, tendon, ligament and nerve
PT tissue growth or regeneration, in wound healing, and in tissue repair and
PT replacement.

XX Claim 20; SEQ ID NO 337; 156pp; English.

XX The invention relates to an isolated polynucleotide encoding a
XX polypeptide with biological activity. The polynucleotides and
XX polypeptides are useful in diagnostics, forensics, gene mapping, and
XX identification of mutations responsible for genetic disorders and other
XX traits, to assess biodiversity, as nutritional sources or supplements.
XX The polynucleotides may also be used as molecular weight markers.
XX chromosome markers or map related gene positions, or as an antigen to
XX raise anti-DNA antibodies or elicit immune response. The polypeptides are
XX useful for raising antibodies, as markers for tissues in which the
XX corresponding polypeptide is expressed, for re-engineering damaged or
XX diseased tissues, for treating myeloid or lymphoid cell disorders, in
XX bone cartilage, tendon, ligament and/or nerve tissue growth or
XX regeneration, in wound healing, in tissue repair and replacement, in
XX healing of burns, incisions and ulcers, and in treating cancer. The
XX present sequence represents the amino acid sequence of a novel human
XX protein.

XX Sequence 395 AA;

Query Match 100.0%; Score 2056; DB 7; Length 395;
Best Local Similarity 100.0%; Pred. No. 3.1e-207;
Matches 395; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MWLPRVSTAVTALLAQTLLFLVSRPGSPAGGEARVHVLVLSMRSGSFGQLF 60
DB 1 MWLPRVSTAVTALLAQTLLFLVSRPGSPAGGEARVHVLVLSMRSGSFGQLF 60
QY 61 NQHPDVFYLMPEPAWHVMTTILSQGSAATLHMAVRDLVRSVFLCDMDVFDAYLPWRRNLSDL 120
DB 61 NQHPDVFYLMPEPAWHVMTTILSQGSAATLHMAVRDLVRSVFLCDMDVFDAYLPWRRNLSDL 120
QY 121 FQWAVSRALCSPACSAFPRGAISSEAVCKPLCARQSFLLAREACRSYSHVVLKEVRFN 180
DB 121 FQWAVSRALCSPACSAFPRGAISSEAVCKPLCARQSFLLAREACRSYSHVVLKEVRFN 180
QY 181 LQVLYPLLSDPALNLRIVHLVRDPRAVLSREQTAKALARDNGIVLGTNGTWVEADPGLR 240
DB 181 LQVLYPLLSDPALNLRIVHLVRDPRAVLSREQTAKALARDNGIVLGTNGTWVEADPGLR 240
QY 241 VREVCRSYHRIAEEAATLKPPPLGRGRYLRVRFEDLAREPLAEIRALYAFTGLSLTPQLE 300
DB 241 VREVCRSYHRIAEEAATLKPPPLGRGRYLRVRFEDLAREPLAEIRALYAFTGLSLTPQLE 300
QY 301 AWIHNTHGSGPGARREAFKTSRNALNVSOAWRHALPFAKIRRVQELCAGALQLLGYRP 360
DB 301 AWIHNTHGSGPGARREAFKTSRNALNVSOAWRHALPFAKIRRVQELCAGALQLLGYRP 360
QY 361 VYSEDEQRNALDLVLPRLNGFTWASSTASHPRN 395
DB 361 VYSEDEQRNALDLVLPRLNGFTWASSTASHPRN 395

RESULT 5
ADL61235

ID ADL61235 standard; protein; 395 AA.

XX ADL61235;

DT 03-JUN-2004 (first entry)

DE Human tyrosine kinase biomarker carbohydrate sulphotransferase 6 protein.
XX predictor set; protein tyrosine kinase; cytostatic; antiangiogenic;
KW vasotrophic; vulnery; pharmacogenomic; drug sensitivity; breast cancer;
KW hypervascular disease; angiogenesis; wound healing scar; human;
KW biomarker; carbohydrate sulphotransferase 6; enzyme.

XX Homo sapiens.
XX WO2004020583-A2.
XX 11-MAR-2004.
XX 26-AUG-2003; 2003WO-US026491.
XX 27-AUG-2002; 2002US-0406385P.
XX (BRIM) BRISTOL-MYERS SQUIBB CO.
PI Huang F, Han X, Reeves KA, Amler L, Fairchild CR, Lee FY;
PI Shaw P;
XX WPI; 2004-239171/22.
DR N-PSDB; ADL61098.
DR New predictor sets with a plurality of polynucleotides and/or
XX polypeptides whose expression pattern predicts cell response to a
PT compound that modulates protein tyrosine kinase activity, useful in
PT treating breast cancer.
XX
PS Claim 9; SEQ ID NO 159; 649pp; English.

XX The invention relates to a novel predictor set comprising a plurality of
XX polynucleotides and/or polypeptides whose expression pattern is
XX predictive of the response of cells to treatment with a compound that
XX modulates protein tyrosine kinase activity or members of the protein
XX tyrosine kinase pathway. The molecules of the invention demonstrate
XX cytostatic, antiangiogenic, vasotrophic and vulnery activities and may
XX be useful in the field of pharmacogenomics, in particular for determining
XX drug sensitivity and in treating breast cancer, hypervascular diseases,
XX angiogenesis and scars in wound healing. The current sequence is that of
XX a human protein tyrosine kinase biomarker protein of the invention.
XX Sequence 395 AA;

Query Match 100.0%; Score 2056; DB 8; Length 395;
Best Local Similarity 100.0%; Pred. No. 3.1e-207;
Matches 395; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MWLPRVSTAVTALLAQTLLFLVSRPGSPAGGEARVHVLVLSMRSGSFGQLF 60
DB 1 MWLPRVSTAVTALLAQTLLFLVSRPGSPAGGEARVHVLVLSMRSGSFGQLF 60
QY 61 NQHPDVFYLMPEPAWHVMTTILSQGSAATLHMAVRDLVRSVFLCDMDVFDAYLPWRRNLSDL 120
DB 61 NQHPDVFYLMPEPAWHVMTTILSQGSAATLHMAVRDLVRSVFLCDMDVFDAYLPWRRNLSDL 120
QY 121 FQWAVSRALCSPACSAFPRGAISSEAVCKPLCARQSFLLAREACRSYSHVVLKEVRFN 180
DB 121 FQWAVSRALCSPACSAFPRGAISSEAVCKPLCARQSFLLAREACRSYSHVVLKEVRFN 180
QY 181 LQVLYPLLSDPALNLRIVHLVRDPRAVLSREQTAKALARDNGIVLGTNGTWVEADPGLR 240
DB 181 LQVLYPLLSDPALNLRIVHLVRDPRAVLSREQTAKALARDNGIVLGTNGTWVEADPGLR 240
QY 241 VREVCRSYHRIAEEAATLKPPPLGRGRYLRVRFEDLAREPLAEIRALYAFTGLSLTPQLE 300
DB 241 VREVCRSYHRIAEEAATLKPPPLGRGRYLRVRFEDLAREPLAEIRALYAFTGLSLTPQLE 300
QY 301 AWIHNTHGSGPGARREAFKTSRNALNVSOAWRHALPFAKIRRVQELCAGALQLLGYRP 360
DB 301 AWIHNTHGSGPGARREAFKTSRNALNVSOAWRHALPFAKIRRVQELCAGALQLLGYRP 360
QY 361 VYSEDEQRNALDLVLPRLNGFTWASSTASHPRN 395
DB 361 VYSEDEQRNALDLVLPRLNGFTWASSTASHPRN 395

RESULT 6

AB081555
ID ABB081555 standard; protein; 395 AA.
XX
AC ABB081555;
XX
DT 05-SEP-2002 (first entry)
XX
DE Consensus N-acetylglucosamine-6-sulfotransferase SEQ ID NO:3.
XX
KW Human; N-acetylglucosamine-6-sulfotransferase; enzyme; GlcNAc6ST;
KW cornal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
KW ophthalmological.
XX
OS Homo sapiens.
OS Mus musculus.
OS Synthetic.
XX
XX
XX Key Location/Qualifiers
FT Misc-difference 10 /label= Ala, Thr, Val
FT Misc-difference 13 /label= Ala, Val, Ser
FT Misc-difference 20 /label= Phe, Cys, Gly
FT Misc-difference 39 /label= Ala, Asp, Glu
FT Misc-difference 96 /label= Val, Met, Ile
FT Misc-difference 142 /label= Ala, Thr, Asn
FT Misc-difference 147 /label= Ala, Asp, Glu
FT Misc-difference 159 /label= Thr, Ser, Gly
FT Misc-difference 238 /label= Gly, His, Arg
FT Misc-difference 294 /label= Ser, Thr, Gly
FT Misc-difference 371 /label= Ala, Thr, Ser
FT Misc-difference 380 /label= Leu, Pro, Met
FT Misc-difference 382 /label= Gly, His, Ser
FT Misc-difference 384 /label= Thr, Ser, Lys
FT Misc-difference 390 /label= Ala, Glu
FT Misc-difference 391 /label= Ser, Lys
FT Misc-difference 392 /label= His, Gln
FT Misc-difference 394 /label= Arg, Glu
FT Misc-difference 395 /label= Asn, Ser
XX
XX US2002061562-A1.
XX
XX 23-MAY-2002.
XX
XX 09-AUG-2001; 2001US-00927602.
XX
XX 11-AUG-2000; 2000US-00838211.
XX
XX 11-AUG-2000; 2000US-0325773P.
XX
XX (FUKU/) FUKUDA M N.
XX (AKAW/) AKAWA T O.
XX
XX Fukuda MN, Akama TO;
XX
XX WPI; 2002-507643/54.
XX

PT New nucleic acid encoding corneal N-acetylglucosamine-6-sulfotransferase,
PT useful for treatment, monitoring and diagnosis of macular corneal
XX dystrophy.
XX
XX Example 5; Fig 2A-B; 69pp; English.
XX
XX The present invention describes human corneal N-acetylglucosamine-6-
XX sulfotransferase (I), which is able to catalyze sulfation of keratan
XX sulfate (KS). Also described is a method for monitoring the effect of
XX treatments for macular corneal dystrophy (MCD), and detecting
XX susceptibility to MCD. (I) is located to chromosome 16q22, and has
XX ophthalmological activity. (I) can be used to treat or prevent macular
XX corneal dystrophy types I or II. (I) makes possible treatment of MCD
XX without requiring keratoplasty or keratectomy. The present sequence
XX represents a consensus N-acetylglucosamine-6-sulfotransferase which is
XX given in the exemplification of the present invention
XX
SQ Sequence 395 AA;

Query Match 94.1%; Score 1934; DB 5; Length 395;
Best Local Similarity 94.4%; Pred. No. 2.2e-194;
Matches 371; Conservative 3; Mismatches 19; Indels 0; Gaps 0;

QY 1 MWLPVSTAVTALLLAQTFLLLFLVSRPGSPSPAGGEARVHVLVSSWSSGSGFVQGLF 60
DB 1 MWLPFSSTXTVXLLLAQTLLFLVSRPGSPSPAGGEXRHHVLVSSWSSGSGFVQGLF 60

QY 61 NQHPDVYLMPEAWHVVTTLSQGSAAATLHMAVRDLVRSVFLCDMDVFDAYLPWRNLSDL 120
DB 61 SQHPDVYLMPEAWHVVTTLSQGSAAATLHMAVRDLVRSVFLCDMDVFDAYLPWRNLSDL 120

QY 121 FQAVSRALCSPSPACSAFPRGAISSBAVKPLCARQSFLLAREACRSYSHVVLKEVRFN 180
DB 121 FQAVSRALCSPSPACSAFPRGXISSEKVKPLCARQFPFLAREACRSYSHVVLKEVRFN 180

QY 181 LQVLYPLSDPALNLRVHVRDPRVAVLSRQETAKALARDNGIVLCTNGTWEADPGLR 240
DB 181 LQVLYPLSDPALNLRVHVRDPRVAVLSRQETAKALARDNGIVLCTNGTWEADPGLR 240

QY 241 VREVCRSVRIAEAAATLKPPLRGYRLVRVEDLAREPLAEIRALYAFYTGSLTPOLE 300
DB 241 VREVCRSVRIAEAAATLKPPLRGYRLVRVEDLAREPLAEIRALYAFYTGSLTPOLE 300

QY 301 AWIHNTHGSGPGARREAFKTSRNALNVSQAWRHALPFAKIRRVQELCAGALQLLYRP 360
DB 301 AWIHNTHGSGPGARREAFKTSRNALNVSQAWRHALPFAKIRRVQELCAGALQLLYRP 360

QY 361 VYSEDEQRLDLVLPGRGXDFXWASSTXXXP 393
DB 361 VYSEDEQRLDLVLPGRGXDFXWASSTXXXP 393

RESULT 7
AA72639
ID AA72639 standard; protein; 390 AA.
XX
XX AA72639;
XX
XX 02-MAY-2001 (first entry)
XX
XX Human glycosyl sulfotransferase-4alpha (GST-4alpha).
XX
XX Human; glycosyl sulfotransferase-4alpha; GST-4alpha; immunosuppressive;
XX therapy; selectin binding inhibitor; gene therapy; inflammation;
XX systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
XX polyarthritis nodosa; polymyositis; systemic sclerosis; dermatitis;
XX glomerulonephritis; myasthenia gravis; Sjogren's syndrome; anemias;
XX Hashimoto's disease; Grave's disease; hypoparathyroidism; anemias;
XX demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
XX myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
XX asthma; hypersensitivity; rheumatic fever; tissue rejection;
XX chromosome 16q23.1.
XX

OS Homo sapiens.
 PN WO200106015-A1.
 XX 25-JAN-2001.
 XX 19-JUL-2000; 2000WO-US019741.
 XX 20-JUL-1999; 99US-0144694P.
 PR 13-JUN-2000; 2000US-00593828.
 XX (REGC) UNIV CALIFORNIA.
 PA Rosen SD, Lee JK, Hemmerich S;
 PI WPI; 2001-138471/14.
 DR N-PSDB; AAD02697, AAD02699, AAD02699.
 XX New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
 PT diagnostic and therapeutic agent screening applications.
 PT Claim 3; Fig 1; 128pp; English.
 PS
 XX The present sequence is human glycosyl sulfotransferase-4alpha (GST-4
 CC alpha). GST-4 gene is found on chromosome 16q23.1. GST is a type 2
 CC membrane protein useful for inhibiting a binding event between a selectin
 CC and a selectin ligand, which comprises contacting the selectin with a non
 CC -sulphated selectin ligand, GST and a small molecular agent that inhibits
 CC the sulphation activity of GST. GST is also useful in inhibiting a
 CC selectin mediated binding event. GST is useful in gene therapy to treat
 CC disorders such as acute or chronic inflammation, systemic lupus
 CC erythematosus (SLE), rheumatoid arthritis, polyarteritis nodosa,
 CC polymyositis, dermatomyositis, systemic sclerosis, diabetes,
 CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
 CC disease, Grave's disease, adrenailitis, hypoparathyroidism, pernicious
 CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
 CC dermatitis, myocarditis, regional enteritis, adult respiratory distress
 CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
 CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
 CC during transplantation. Note: The present sequence is also shown in
 CC sequence listing (page no: 56) but lacks four nucleotides at its 3' end
 XX
 SQ Sequence 390 AA;
 Query Match 84.1%; Score 1729.5; DB 4; Length 390;
 Best Local Similarity 85.8%; Pred. No. 7.5e-173; Mismatches 17; Indels 1; Gaps 1;
 Matches 333; Conservative 17; Mismatches 37; Indels 1; Gaps 1;
 QY 1 MWLPRVSVSTAVTALLAQ-TFLLFLVSRPGPSPAGGEARVHVLVLSWSRSGSFVGLQ 59
 DB 1 MWLPRSSKTVTVLLLAQTTCLLFIISRPGPSPAGGEDRVHVLVLSWSRSGSFVGLQ 60
 QY 60 FNOHPDVFYLMPEPAWHVWTTLSQGSAAATHMAVRDLVRSVFLCDMDVFDAYLPWRNLS 119
 DB 61 FSQHPDVFYLMPEPAWHVWTTLSQGSAAATHMAVRDLVRSVFLCDMDVFDAYMPQSNLSA 120
 QY 120 LFQWAVSRALCSPACSPAGGAPRGAISSEAVCKPLCARQSFLLAREACRSYSHVVLKEVRFF 179
 DB 121 FFWNATSRALCSPACSPAGGAPRGTISQDVCKTQTRQPSFLAREACRSYSHVVLKEVRFF 180
 QY 180 NLQVLYLLSDPALNLRIVHLVRDPRAVLRSRQETAKALARDNGIVLGTNGTWEADPGL 239
 DB 181 NLQVLYLLSDPALNLRIVHLVRDPRAVLRSRQETAKALARDNGIVLGTNGTWEADPGL 240
 QY 240 RVVRECRSHVRAEAATLKPPPLKGRVRLVRFEDLAREPLAEIRALYAFGLSLTPQL 299
 DB 241 RLIREVCRSHVRAEAATLKPPPLKGRVRLVRFEDLAREPLAEIRALYAFGLSLTPQL 300
 QY 300 EAMHNITHGSGPGARREAPKTSRRNALNVQAWRHLPKAKIRRYQOELCAGALQLGYR 359
 DB 301 EAMHNITHGSGGKPIEAPHTSSRNARNVQAWRHLPKAKIRRYQOELCAGALQLGYR 360
 QY 360 PVYSEDEQRNALDLVLPRGLNGFTWAS 387

DB 361 PVYSADQQRDLTDLVLRGPDHFSWAS 388
 ||||| :||:| :||||| :|:|
 RESULT 8.
 ABB81556
 ID ABB81556 standard; protein; 390 AA.
 XX ABB81556;
 AC ABB81556;
 XX 05-SEP-2002 (first entry)
 DT Human intestinal N-acetylglucosamine-6-sulfotransferase SEQ ID NO:4.
 DE Human; N-acetylglucosamine-6-sulfotransferase; enzyme; GlcNAc6ST;
 KW corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
 KW ophthalmological.
 XX Homo sapiens.
 OS US2002061562-A1.
 XX 23-MAY-2002.
 PD 09-AUG-2001; 2001US-00927602.
 XX 11-AUG-2000; 2000US-00638211.
 PR 11-AUG-2000; 2000US-0325773P.
 XX (FUKU/) FUKUDA M N.
 PA (AKAM/) AKAMA T O.
 PI Fukuda MN, Akama TO;
 XX WPI; 2002-507643/54.
 DR New nucleic acid encoding corneal N-acetylglucosamine-6-sulfotransferase,
 XX useful for treatment, monitoring and diagnosis of macular corneal
 PT dystrophy.
 XX Example 5; Fig 2A-B; 69pp; English.
 CC The present invention describes human corneal N-acetylglucosamine-6-
 CC sulfotransferase (I), which is able to catalyse sulfation of keratan
 CC sulfate (KS). Also described is a method for monitoring the effect of
 CC treatments for macular corneal dystrophy (MCD), and detecting
 CC susceptibility to MCD. (I) is located to chromosome 16q22, and has
 CC ophthalmological activity. (I) can be used to treat or prevent macular
 CC corneal dystrophy types I or II. (I) makes possible treatment of MCD
 CC without requiring keratinoplasty or keratectomy. The present sequence
 CC represents human intestinal N-acetylglucosamine-6- sulfotransferase,
 CC which is given in comparison with (I) in the exemplification of the
 CC present invention
 XX
 SQ Sequence 390 AA;
 Query Match 84.1%; Score 1729.5; DB 5; Length 390;
 Best Local Similarity 85.8%; Pred. No. 7.5e-173; Mismatches 17; Indels 1; Gaps 1;
 Matches 333; Conservative 17; Mismatches 37; Indels 1; Gaps 1;
 QY 1 MWLPRVSVSTAVTALLAQ-TFLLFLVSRPGPSPAGGEARVHVLVLSWSRSGSFVGLQ 59
 DB 1 MWLPRSSKTVTVLLLAQTTCLLFIISRPGPSPAGGEDRVHVLVLSWSRSGSFVGLQ 60
 QY 60 FNOHPDVFYLMPEPAWHVWTTLSQGSAAATHMAVRDLVRSVFLCDMDVFDAYLPWRNLS 119
 DB 61 FSQHPDVFYLMPEPAWHVWTTLSQGSAAATHMAVRDLVRSVFLCDMDVFDAYMPQSNLSA 120
 QY 120 LFQWAVSRALCSPACSPAGGAPRGAISSEAVCKPLCARQSFLLAREACRSYSHVVLKEVRFF 179
 DB 121 FFWNATSRALCSPACSPAGGAPRGTISQDVCKTQTRQPSFLAREACRSYSHVVLKEVRFF 180
 QY 180 NLQVLYLLSDPALNLRIVHLVRDPRAVLRSRQETAKALARDNGIVLGTNGTWEADPGL 239

Db 181 NLOVYLLSDPALNLRIVHLVRDPRVAVLSREAAGPILARDNGIVLTGKWEADPHL 240
 Qy 240 RVREVCRSVHRTAEATLKPPLGRYRLVRFDIAREPLAEIRALYAFGLTLPOL 299
 Db 241 RLIREVCRSVHRTAEATLKPPLGRYRLVRFDIAREPLAEIRALYAFGLTLPOL 300
 Qy 300 EAMHNTHSGGPGARREAFKTSRNALNSQAWRHLPFAKIRRVQELCAGALQLGYR 359
 Db 301 EAMHNTHSGGPGARREAFKTSRNALNSQAWRHLPFAKIRRVQELCAGALQLGYR 360
 Qy 360 PVYSEDEQRNALDLVLPRLGNGFTWAS 387
 Db 361 PVYSAQDQRLDLVLPRLGPDHFSWAS 388

RESULT 9

AAV72638
 ID AAV72638 standard; protein; 395 AA.

XX
 AC AAV72638;

XX
 DT 02-MAY-2001 (first entry)

XX
 DE Mouse glycosyl sulfotransferase-4 (GST-4).

XX
 KW Mouse; glycosyl sulfotransferase-4; GST-4; immunosuppressive; therapy;
 KW selectin binding inhibitor; gene therapy; inflammation;
 KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
 KW polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
 KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenalitis;
 KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
 KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
 KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
 KW asthma; hypersensitivity; rheumatic fever; tissue rejection;
 KW chromosome 8E1.

XX
 OS Mus musculus.

XX
 PN W0200106015-A1.

XX
 PD 25-JAN-2001.

XX
 PF 19-JUL-2000; 2000WO-US019741.

XX
 PR 20-JUL-1999; 99US-0144694P.

XX
 PR 13-JUN-2000; 2000US-00593828.

XX
 PA (REGC) UNIV CALIFORNIA.

XX
 PI Rosen SD, Lee JK, Hammerich S;

XX
 DR WPI; 2001-138471/14.

XX
 DR N-PSDB; AAD02696.

PT New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
 PT diagnostic and therapeutic agent screening applications.

XX
 PS Claim 3; Fig 2; 128pp; English.

XX
 CC The present sequence is mouse glycosyl sulfotransferase-4 (GST-4). GST-4
 CC gene is found on chromosome 8E1. GST is a type 2 membrane protein useful
 CC for inhibiting a binding event between a selectin and a selectin ligand,
 CC which comprises contacting the selectin with a non-sulphated selectin
 CC ligand, GST and a small molecular agent that inhibits the sulphation
 CC activity of GST. GST is also useful in inhibiting a selectin mediated
 CC binding event. GST is useful in gene therapy to treat disorders such as
 CC acute or chronic inflammation, systemic lupus erythematosus (SLE),
 CC rheumatoid arthritis, polyarteritis nodosa, polymyositis,
 CC dermatomyositis, systemic sclerosis, diabetes, glomerulonephritis,
 CC myasthenia gravis, Sjogren's syndrome, Hashimoto's disease, Grave's
 CC disease, adrenalitis, hypoparathyroidism, pernicious anaemia,
 CC demyelinating diseases, cirrhosis, ulcerative colitis, dermatitis,

CC myocarditis, regional enteritis, adult respiratory distress syndrome,
 CC infantile eczema, psoriasis lichen planus, allergic rhinitis, bronchial
 CC asthma, hypersensitivity, rheumatic fever and tissue rejection during
 CC transplantation
 XX Sequence 395 AA;

Query Match 83.1%; Score 1708; DB 4; Length 395;

Best Local Similarity 83.0%; Pred. No. 1.4e-170;

Matches 328; Conservative 24; Mismatches 43; Indels 0; Gaps 0;

Qy 1 MMLPRVSTAVTALLAQTLLFLVSRPGPSPAGAEARVHVLVLSWSSGSSFVGOLF 60

Db 1 MRLPRFSSTVMLSLMVQTLVFLVSRQVPSPAGLGERVHVLVLSWSSGSSFVGOLF 60

Qy 61 NOHPDVFLMEPAWHVYVTTLSQGSAAATLHMAVRDLVRSVFLCDMDVFDAYLPHRRNLSDL 120

Db 61 SQHPDVFLMEPAWHVYVTTLSQGSAPALHMAVRDLVRSVFLCDMDVFDAYLPHRRNLSDL 120

Qy 121 FQWAVSRALCSPACSAFPRGAISSEAVCKPLCARQSFITLAREACRSYSHVVLKEVRFN 180

Db 121 FQWAVSRALCSPVCEAFARGNISSEVCKPLCATRPFGLAQACSSYSHVVLKEVRFN 180

Qy 181 LQVLYPLSDPALNLRIVHLVRDPRVAVLSRREQTAKALARDNGIVLTGNGTWVEADPLR 240

Db 181 LQVLYPLSDPALNLRIVHLVRDPRVAVLSRREQTAKALARDNGIVLTGNGTWVEADPLR 240

Qy 241 VREVCRSVHRTAEATLKPPLGRYRLVRFDIAREPLAEIRALYAFGLTLPOL 300

Db 241 VNEVCRSVHRTAEALHPPPLQDRLVRYVEDLARDPLTVIRELYATGTLTLPOL 300

Qy 301 AWIHNTHSGGPGARREAFKTSRNALNSQAWRHLPFAKIRRVQELCAGALQLGYR 360

Db 301 TWIHNTHSGGPGARREAFKTSRDALSQAWRHLPFAKIRRVQELCGALQLGYRS 360

Qy 361 VYSEDEQRNALDLVLPRLGNGFTWASSTASHPRN 395

Db 361 VHSELEQRDLSDLLPLRGMDSPFKWASSTEKQPS 395

RESULT 10

AAU11275

ID AAU11275 standard; protein; 395 AA.

XX
 AC AAU11275;

XX
 DT 12-MAR-2002 (first entry)

XX
 DE Murine intestinal-GlcNAc 6-sulfotransferase (I-GlcNAc6ST) protein.

XX
 KW Mouse; beta1,3GNT; beta1,3-N-acetylglucosaminyltransferase; MECA-79;
 KW L-selectin; L-selectin sulfotransferase-2; Crohn's disease; diabetes;
 KW ulcerative colitis; inflammatory skin disorder; psoriasis; lichen planus;
 KW allergic contact dermatitis; lymphoma; chronic pneumonia; USS-2;
 KW delayed-type hypersensitivity reaction; hyperplastic thymus; antileuk;
 KW antiinflammatory; antipsoriatic; antidiabetic; dermatological;
 KW antiallergic; intestinal-GlcNAc 6-sulfotransferase; I-GlcNAc6ST.

XX
 OS Mus musculus.

XX
 PN W0200185177-A1.

XX
 PD 15-NOV-2001.

XX
 PF 10-MAY-2001; 2001WO-US015452.

XX
 PR 11-MAY-2000; 2000US-00569320.

XX
 PA (BURN-) BURNHAM INST.

XX
 PI Fukuda M, Yeh J, Hiraoka N;

XX
 DR WPI; 2002-075226/10.

DR N-PSDB; AAS16948.
XX New enzyme, useful for modifying acceptor molecule, comprises an isolated
PT L-selectin sulfotransferase-2 that directs expression of L-selectin
PT ligand antigen, MECA-79 in Chinese hamster ovary cells, or intestinal
PT GlcNAc 6-sulfotransferase.
XX
PS Claim 28; Fig 10; 98pp; English.
XX
XX The present invention provides a method of modifying an acceptor molecule
CC by contacting the acceptor with an isolated beta1,3-N-
CC acetylglucosaminyltransferase (beta1,3Gnt) or an active fragment, where
CC beta1,3Gnt directs expression of a MECA-79 antigen. The invention also
CC provides a method of treating or preventing an L-selectin-mediated
CC condition by reducing the expression or activity of a beta1,3Gnt that
CC directs expression of a MECA-79 antigen. This can be done by
CC administering to the subject an oligosaccharide L-selectin antagonist
CC that inhibits binding of L-selectin to a MECA-79 antigen, for example by
CC administering antibody material that specifically binds beta1,3Gnt,
CC and/or a beta1,3Gnt antisense nucleic acid molecule. L-selectin
CC sulfotransferase-2 (LSST-2) also directs MECA-79 antigen expression.
CC Alternatively, the expression or activity of LSST-2 or its active
CC fragment can be reduced in combination with reducing the expression or
CC activity of beta1,3Gnt. The method is useful for treating L-selectin
CC mediated conditions such as Crohn's disease and ulcerative colitis,
CC inflammatory disorders of the skin such as allergic contact dermatitis,
CC psoriasis and lichen planus, lymphomas, chronic pneumonia, delayed-type
CC hypersensitivity reactions, diabetes and hyperplastic thymus. This
CC sequence represents mouse I-GlcNAc6ST
XX
SQ Sequence 395 AA;

Query Match 83.1%; Score 1708; DB 5; Length 395;
Best Local Similarity 83.0%; Pred. No. 1.4e-170; Indels 0; Gaps 0;
Matches 328; Conservative 24; Mismatches 43; Indels 0; Gaps 0;
QY 1 MWLPRVSTAVTALLAQTLLFLVSRPGSPAGGEARVHVLVLSWRSGSSFVGQLF 60
DB 1 MRLPRFSSTVMSLLMVQTGILVFLVSRQVPSPAGLGERVHVLVLSWRSGSSFVGQLF 60
QY 61 NQHPDVFYLMPEPAWHVMTTLSQGSAAATLHMAVRDLVRSVFLCDMDVFDAYLPWRRNLSL 120
DB 61 SQHPDVFYLMPEPAWHVMTTLSQGSAPALHMAVRDLIRSVFLCDMDVFDAYLPWRRNLSL 120
QY 121 FQWAVSRALCSPACSAFPRGAISSAEVCKPLCARQSFLLAREACRSYSHVVLKEVRFFN 180
DB 121 FQWAVSRALCSPVCEAFARGNISSEVCKPLCATRPFGLAQEACSSYSHVVLKEVRFFN 180
QY 181 LQVLYPLLSDPALNLRIVHLVRDPRVLRSRQTKALARDNGIVLGTNGTWVEADPGLR 240
DB 181 LQVLYPLLSDPALNLRIVHLVRDPRVLRSRQTKALARDNGIVLGTNGTWVEADPGLR 240
QY 241 VVREVCRRSHVRIAEATLKPPPLRGYRLVRFPEDLAREPLAEIRALYFTGLSLTPQLE 300
DB 241 VVNEVCRRSHVRIAEAAHLKPPPLQDRYRLVRVEDLARDPLTVIRELYAFTGLSLTPQLE 300
QY 301 AWHIHTHSGGPGARREAFKTSRRNALNVSOAWRHLPFAKIRRVQELCAGALQLLGYRP 360
DB 301 TWIHNHTHSGGPGARREAFKTSRDALSVSQAWRHLPFAKIRRVQELCGGALQLLGYRS 360
QY 361 VYSEDEQRNALDLVLRGLNGFTWASSTASHPRN 395
DB 361 VHSELRQDLSDLLLPGRGNSDFKMASSTEKQES 395

RESULT 11
ABB81557
ID ABB81557 standard; protein; 418 AA.

XX ABB81557;

XX 05-SEP-2002 (first entry)

XX

DE Mouse intestinal N-acetylglucosamine-6-sulfotransferase SEQ ID NO:5.
XX
KW Human; N-acetylglucosamine-6-sulfotransferase; enzyme; GlcNAc6ST;
KW corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
KW ophthalmological.
XX
OS Mus musculus.
XX
PN US2002061562-A1.
XX
PD 23-MAY-2002.
XX
PF 09-AUG-2001; 2001US-00927602.
XX
XX 11-AUG-2000; 2000US-00638211.
PR 11-AUG-2000; 2000US-0325773P.
XX
XX (FUKU/) FUKUDA M N.
PA (AKAM/) AKAMA T O.
XX
PI Fukuda MN, Akama TO;
XX
XX WPI; 2002-507643/54.
DR
XX New nucleic acid encoding corneal N-acetylglucosamine-6-sulfotransferase,
PT useful for treatment, monitoring and diagnosis of macular corneal
PT dystrophy.
XX
XX Example 5; Page 24-25; 69pp; English.
XX
XX The present invention describes human corneal N-acetylglucosamine-6-
CC sulfotransferase (I), which is able to catalyze sulfation of keratan
CC sulfate (KS). Also described is a method for monitoring the effect of
CC treatments for macular corneal dystrophy (MCD), and detecting
CC susceptibility to MCD. (I) is located to chromosome 16q22, and has
CC ophthalmological activity. (I) can be used to treat or prevent macular
CC corneal dystrophy types I or II. (I) makes possible treatment of MCD
CC without requiring keratinoplasty or keratectomy. The present sequence
CC represents mouse intestinal N-acetylglucosamine-6- sulfotransferase,
CC which is given in comparison with (I) in the exemplification of the
CC present invention
XX
SQ Sequence 418 AA;

Query Match 82.4%; Score 1694.5; DB 5; Length 418;
Best Local Similarity 82.8%; Pred. No. 4.1e-169; Indels 1; Gaps 1;
Matches 327; Conservative 24; Mismatches 43; Indels 1; Gaps 1;
QY 1 MWLPRVSTAVTALLAQTLLFLVSRPGSPAGGEARVHVLVLSWRSGSSFVGQLF 60
DB 25 MRLPRFSSTVMSLLMVQTGILVFLVSRQVPSPAGLGERVHVLVLSWRSGSSFVGQLF 84
QY 61 NQHPDVFYLMPEPAWHVMTTLSQGSAAATLHMAVRDLVRSVFLCDMDVFDAYLPWRRNLSL 120
DB 85 SQHPDVFYLMPEPAWHVMTTLSQGSAPALHMAVRDLIRSVFLCDMDVFDAYLPWRRNLSL 144
QY 121 FQWAVSRALCSPACSAFPRGAISSAEVCKPLCARQSFLLAREACRSYSHVVLKEVRFFN 180
DB 145 FQWAVSRALCSPVCEAFARGNISSEVCKPLCATRPFGLAQEACSSYSHVVLKEVRFFN 204
QY 181 LQVLYPLLSDPALNLRIVHLVRDPRVLRSRQTKALARDNGIVLGTNGTWVEADPGLR 240
DB 205 LQVLYPLLSDPALNLRIVHLVRDPRVLRSRQTKALARDNGIVLGTNGTWVEADPGLR 264
QY 241 VVREVCRRSHVRIAEATLKPPPLRGYRLVRFPEDLAREPLAEIRALYFTGLSLTPQLE 300
DB 265 VVNEVCRRSHVRIAEALH-KPPPLQDRYRLVRVEDLARDPLTVIRELYAFTGLSLTPQLE 323
QY 301 AWHIHTHSGGPGARREAFKTSRRNALNVSOAWRHLPFAKIRRVQELCAGALQLLGYRP 360
DB 324 TWIHNHTHSGGPGARREAFKTSRDALSVSQAWRHLPFAKIRRVQELCGGALQLLGYRS 383
QY 361 VYSEDEQRNALDLVLRGLNGFTWASSTASHPRN 395

DB 384 VHSEQRDLSDLLPRGMDSKFASSTKQPS 418
 RESULT 12
 AAB41947
 ID AAB41947 standard; protein; 418 AA.
 XX AAB41947;
 XX 08-FEB-2001 (first entry)
 XX Human ORFX ORF1711 polypeptide sequence SEQ ID NO:3422.
 DE
 XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 XX vulnary; antipsoriatic; antiparkinsonian; nontropic; neuroprotective;
 XX anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 XX immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 XX hypotensive; dermatological; immunosuppressive; antiinflammatory;
 XX antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 XX antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
 XX neurodegenerative disorder; osteoarthritis; graft vs host disease;
 XX cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 XX cholesterol ester storage; systemic lupus erythematosus; infection;
 XX severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 XX allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 XX bone damage; cartilage damage; antiinflammatory disease; coagulation;
 XX thrombosis; contraceptive.
 XX Homo sapiens.
 OS
 XX WO200058473-A2.
 PN
 XX 05-OCT-2000.
 PD
 XX 31-MAR-2000; 2000WO-US008621.
 PF
 XX 31-MAR-1999; 99US-0127607P.
 PR 02-APR-1999; 99US-0127636P.
 PR 05-APR-1999; 99US-0127728P.
 PR 30-MAR-2000; 2000US-00540763.
 XX (CURA-) CURAGEN CORP.
 XX Shimkets RA, Leach M;
 XX WPI; 2000-602362/57.
 DR N-PSDB; AAC76156.
 XX Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease.
 XX Claim 11; Page 2599-2600; 5507pp; English.
 PS
 XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
 CC antipsoriatic; antiparkinsonian; nontropic; neuroprotective; osteopathic;
 CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
 CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
 CC dermatological; immunosuppressive; antiinflammatory; antibacterial;
 CC antiviral; antifungal; antirheumatic; antithyroid; and antianaemic. The
 CC sequences can be used for determining the presence of or predisposition
 CC to, or preventing or treating pathological conditions associated with an
 CC ORFX-associated disorder. The nucleic acids can be used to express ORFX
 CC proteins in gene therapy vectors. The proteins and nucleic acids may be
 CC used to treat cancers, proliferative disorders, neurodegenerative
 CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
 CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
 CC storage, systemic lupus erythematosus, severe combined immunodeficiency
 CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
 CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and

CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
 CC enhance coagulation; to inhibit thrombosis; and as a contraceptive
 XX
 SQ Sequence 418 AA;
 Query Match 78.8%; Score 1620; DB 3; Length 418;
 Best Local Similarity 81.0%; Pred. No. 2.9e-161;
 Matches 315; Conservative 22; Mismatches 50; Indels 2; Gaps 2;
 QY 1 MWLPVYSTAVTALLAQ-TFLLLFLVSRPGSPAGGEARVHVLLVSSWRSGSSFFVQGL 59
 DB 28 MWLPRESSKTVTVLLAQTTCLLLFIISRPSPAGGEDRVHVLLVSSWRSGSSFFVQGL 87
 QY 60 FNOHPDVFYLMPEPAWHVTTLSQSSAATLHMAVRDLVRSVFLCOMDVDFDAYL-PWRNLS 118
 DB 88 FSQHPDVFYLMPEPAWHVTTLSQSSAATLHMAVRDLVRSVFLCOMDVDFDAYMEPGRQPS 147
 QY 119 DLFQWAVSRALCSPACSPACSPACSPACSPACSPACSPACSPACSPACSPACSPACSPAC 178
 DB 148 SLFQWENSRALCSPACSPACSPACSPACSPACSPACSPACSPACSPACSPACSPACSPAC 207
 QY 179 FNLQVLYPLLSDDPALNLRIVHLVRDPAVLRSREQTAKALARDNGIVLGTNGTWVEADPG 238
 DB 208 FNLQVLYPLLSDDPALNLRIVHLVRDPAVLRSREQTAKALARDNGIVLGTNGTWVEADPG 267
 QY 239 LRVVREVCVSHVRIAEAAATLKPPFLGRVRLVRFEDLAREPLAEIRALYAFTGLSTPQ 298
 DB 268 LRLIREVCVSHVRIAEAAATLKPPFLGRVRLVRFEDLAREPLAEIRALYAFTGLSTPQ 327
 QY 299 LEAMINITHGSGFGARREAFKTSRNALNVSOAHRHALPFKIRRVQELCAGALQLLGY 358
 DB 328 LEAMINITHGSGIGKPIEAFTTSRNARNVSOAHRHALPFTKILRVQEVCAAGALQLLGY 387
 QY 359 RPVYSEDEORNTALDLVLPRLNGFTWAS 387
 DB 388 RPVYADQQRDLTLVLPRLNGFTWAS 416
 RESULT 13
 ADJ70405
 ID ADJ70405 standard; protein; 394 AA.
 XX ADJ70405;
 XX 06-MAY-2004 (first entry)
 XX Human heat mitochondrial protein as a therapeutic target SeqID2211.
 XX mitochondrial; human; screening assay; diabetes mellitus;
 XX Huntington's disease; osteoarthritis;
 XX Leber's hereditary optic neuropathy; LHON;
 XX mitochondrial encephalopathy lactic acidosis and stroke; MELAS;
 XX myoclonic epilepsy ragged red fibre syndrome; MERRF; cancer;
 XX neuroprotective; nontropic; antidiabetic; anticonvulsant; antiarthritic;
 XX osteopathic; ophthalmological; cytostatic.
 OS Homo sapiens.
 XX WO2003087768-A2.
 PN 23-OCT-2003.
 PD 04-APR-2003; 2003WO-US010870.
 XX 12-APR-2002; 2002US-0372843P.
 PR 17-JUN-2002; 2002US-0389987P.
 PR 20-SEP-2002; 2002US-0412418P.
 XX (MITO-) MITOKOR.
 PA (BUCK-) BUCK INST AGE RES.
 XX Ghosh SS, Fahy ED, Zhang B, Gibson BW, Taylor SW, Glenn GM;
 PI Warnock DE;

XX WPI; 2003-845369/78.
 XX Identifying a mitochondrial target for drug screening assays and for
 PT treating diseases associated with altered mitochondrial function,
 PT comprises detecting a modified polypeptide in a sample and correlating
 PT with the disease.
 XX
 XX Claim 1; SEQ ID NO 2211; 180pp; English.
 XX
 CC This invention relates to novel mitochondrial targets that can be used
 CC for therapeutic intervention in treating a disease associated with
 CC altered mitochondrial function. Specifically, it refers to a method for
 CC identifying proteins of the human heart mitochondrial proteome that are
 CC useful for drug screening assays, as well as therapeutic targets. The
 CC present invention describes a method for identifying such proteins that
 CC can be used in the treatment of various diseases associated with altered
 CC mitochondrial function including diabetes mellitus, Huntington's disease,
 CC osteoarthritis, Leber's hereditary optic neuropathy (LHON), mitochondrial
 CC encephalopathy lactic acidosis and stroke (MELAS), myoclonic epilepsy
 CC ragged red fibre syndrome (MERRF) or cancer. Accordingly, these
 CC compositions have neuroprotective, neurotropic, antidiabetic,
 CC anticonvulsant, antiarthritic, osteopathic, ophthalmological and
 CC cytoskeletal activities. This polypeptide sequence is a human heart
 CC mitochondrial protein of the invention.
 XX
 XX Sequence 394 AA;

Query Match 74.8%; Score 1538; DB 7; Length 394;
 Best Local Similarity 93.3%; Pred. No. 1.1e-152;
 Matches 294; Conservative 7; Mismatches 14; Indels 0; Gaps 0;

QY 1 MWLPRVSSTAVTALLAQTLLFLVSRPGSPAGGEARVHVLVSSWRSQSSFGVOLF 60
 Db 1 MWLPRVSSTAVTALLAQTLLFLVSRPGSPAGGEARVHVLVSSWRSQSSFGVOLF 60
 QY 61 NQHPDVFYLMPEPAWHVMTTLLSQGSAATLHMAVRDLVRSVFLCDMDVFDAYLPWRRNLSDL 120
 Db 61 NQHPDVFYLMPEPAWHVMTTLLSQGSAATLHMAVRDLVRSVFLCDMDVFDAYLPWRRNLSDL 120
 QY 121 FQWAVSRALCSPACSAFPFGAISSEAVCKPLCARQSFLLAREACRSYSHVVLKEVRF 180
 Db 121 FQWAVSRALCSPACSAFPFGAISSEAVCKPLCARQSFLLAREACRSYSHVVLKEVRF 180
 QY 181 LQVLYPLSDPALNLRIVLHVRDPRVLRSEQTAKALARDNGIVLGTNGTWVEADPGLR 240
 Db 181 LQVLYPLSDPALNLRIVLHVRDPRVLRSEQTAKALARDNGIVLGTNGTWVEADPGLR 240
 QY 241 VREVCRSYHRIAEEATLKPPPLRGYRLVRFPEDLAREPLAIEALYFTGLSLTPQLE 300
 Db 241 LIREVCRSYHRIAEEATLKPPPLRGYRLVRFPEDLAREPLAIEALYFTGLSLTPQLE 300
 QY 301 AWIHNTHGSGPGAR 315
 Db 301 AWIHNTHGSGASQ 315

RESULT 14
 AAM93309
 ID AAM93309 standard; protein; 386 AA.

XX AAM93309;
 AC AAM93309;
 XX
 DT 06-NOV-2001 (first entry)
 XX Human polypeptide, SEQ ID NO: 2817.
 XX Human; full length cDNA; cDNA synthesis; oligo-capping.
 XX Homo sapiens.
 OS
 XX
 XX EPI130094-A2.
 XX

PD 05-SEP-2001.
 XX
 PF 07-JUL-2000; 2000EP-00114089.
 XX
 PR 08-JUL-1999; 99JP-00194486.
 PR 11-JAN-2000; 2000JP-00118774.
 PR 02-MAY-2000; 2000JP-00183765.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 XX
 DR WPI; 2001-524255/58.
 DR N-PSDB; AAK94229.
 XX
 XX 830 Primers useful for synthesizing full length cDNA clones and their use
 PT in genetic manipulation.
 PT
 PS Claim 8; SEQ ID NO 2817; 1380pp + Sequence Listing; English.
 CC The invention relates to primers for synthesising full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been isolated
 CC and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have
 CC been determined. Primers for synthesising the full length cDNA are useful
 CC for clarifying the function of the protein encoded by the cDNA. The full
 CC length clones were obtained by construction of full length enriched cDNA
 CC libraries that were synthesised by the oligo-capping method. The primers
 CC enable the production of the full length cDNA easily without any special
 CC methods. The present sequence is a polypeptide encoded by a full length
 CC human cDNA of the invention. Note: The sequence data for this patent did
 CC not form part of the printed specification, but was obtained in CD-ROM
 CC format directly from EPO
 XX
 XX Sequence 386 AA;

Query Match 49.8%; Score 1024; DB 4; Length 386;
 Best Local Similarity 54.8%; Pred. No. 1.6e-98;
 Matches 217; Conservative 44; Mismatches 107; Indels 28; Gaps 7;

QY 14 LLLAQTFLLLFLVSRP-----GSSSPAGGEARVHVLVSSWRSQSSFGVOLF 59
 Db 2 LLLPKMKLLFLVSRP-----GSSSPAGGEARVHVLVSSWRSQSSFGVOLF 61
 QY 60 FQHPDVFYLMPEPAWHVMTTLLSQGSAATLHMAVRDLVRSVFLCDMDVFDAYLPWRRNLS 118
 Db 62 FQHPDVFYLMPEPAWHVMTTLLSQGSAATLHMAVRDLVRSVFLCDMDVFDAYLPWRRNLS 121
 QY 119 DLFWAVSRALCSPACSAFPFGAISSEAVCKPLCARQSFLLAREACRSYSHVVLKEVRF 178
 Db 122 SLFWAVSRALCSPACSAFPFGAISSEAVCKPLCARQSFLLAREACRSYSHVVLKEVRF 181
 QY 179 FNLQVLYPLSDPALNLRIVLHVRDPRVLRSEQTAKALARDNGIVLGTNGTWVEADP 237
 Db 182 FNLQVLYPLSDPALNLRIVLHVRDPRVLRSEQTAKALARDNGIVLGTNGTWVEADP 241
 QY 238 GLRVVREVCRSYHRIAEEATLKPPPLRGYRLVRFPEDLAREPLAIEALYFTGLSLT 296
 Db 242 PYVNVQVICSQLEIYK--TIQSLPKALQERYLLVRYEDLAREPVAQTSMRYEFVGLLEF 299
 QY 297 POLEAWIHNTHGSGPGARREAFKTSRRNALNVSAWRHALPFKIRRVQELCAGALQLL 356
 Db 300 PHLQVWVHNITRGKMG--DHAFTNARDALNVSAWRWSLPYKVSRLQKACGDAMNLL 357
 QY 357 GYRPVYSEDEQRNALDLVLPRLGNGFTWASSTASH 392
 Db 358 GYRHVRSEQRNLLLDLL-----STWVPEQIH 386

RESULT 15
 ADL30784
 ID ADL30784 standard; protein; 386 AA.
 XX

AC ADL30784;
XX
DT 20-MAY-2004 (first entry)
XX
DE Human protein encoded by a full length cDNA clone SeqID 2817.
XX
KW human; medicine; signal transduction; glycoprotein; transcription;
KW oligo-capping method.
XX
OS Homo sapiens.
XX
PN EF136543-A2.
XX
PD 10-MAR-2004.
XX
PF 07-JUL-2000; 2003EP-00025638.
XX
PR 08-JUL-1999; 99JP-00194486.
PR 11-JAN-2000; 2000JP-00118774.
PR 02-MAY-2000; 2000JP-00183865.
PR 07-JUL-2000; 2000EP-00114089.
XX
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
DR WPI; 2004-204755/20.
DR N-PSDB; ADL30783.
XX
XX New oligonucleotide primers (830 cDNAs) useful for synthesizing full
PT length human cDNAs.
XX
XX Example 1; SEQ ID NO 2817; 1340pp; English.
XX
CC This invention relates to a novel primers useful for synthesizing full
CC length cDNA molecules that encode human proteins. Specifically, it refers
CC to secretory or membrane proteins that are potential therapeutic agents/
CC target molecules in the field of medicine, and in particular genes
CC encoding proteins that are associated with signal transduction,
CC glycoproteins and transcription. The present invention describes a method
CC for efficiently cloning a full length human cDNA from both the 5' and 3',
CC ends using the oligo-capping method. This polypeptide sequence is a full
CC length human protein of the invention.
XX
SQ Sequence 386 AA;

Query Match 49.8%; Score 1024; DB 8; Length 386;
Best Local Similarity 54.8%; Pred. No. 1.6e-98;
Matches 217; Conservative 44; Mismatches 107; Indels 28; Gaps 7;

Qy 14 LLLAQTFLLLFLVSRP-----GPSPGAGEARVHVVLSSWRSGSFVQQL 59
Db 2 LLPKMKLLFLVSQLAILAFPHMYGHNISSLSMKAPERMHVVLSSWRSGSFVQQL 61
Qy 60 FNOHPDVFYLMPEPAWHVYMTLISQGSAAATLHMAVRDLVRSVFLCDMDVFDAYL-PWRNLS 118
Db 62 FGQHPDVFYLMPEPAWHVYMTFTKQTAWMLHMAVRDLIRAVFLCDMSVFDAYMEPCPRQS 121
Qy 119 DLQWAVSRALCSPAPCPGAFPRGAISSEAVCKPLCARQSFTLAREACRSYSHVVLKEVRF 178
Db 122 SLQWENSRALCSAPACDIIIPQDEIIIPRAHCRLLCSQPPFEVVEKACRSYSHVVLKEVRF 181
Qy 179 FNLQVLYPLSDPALNIRIVHLVDRDPAVLRSRQTAKALARDNGIVLGTNGTWV-EADP 237
Db 182 FNLQSLYPLLDPSLNHLHVLVRDPAVFRSRETRKGLMDINRIVMGQHEQKLKXEDQ 241
Qy 238 GLRVVRECRSHVRIAEATLKP-PPFLRGYRLVRFDLAREPLAEIRALYAFGLSLT 296
Db 242 PYYVMQICQSLQEIYK--TIQSLPKALQERYLLVRFDLAREPAVQTSRWYFVGLLEFL 299
Qy 297 POLBAWIHNTHGSGPGGARREAFKTSRNALNVSQAWRHALPPFAKIRRVQELCAGALQLL 356

Db 300 PHLQTVVHNITRGKMG--DHAFHTNARDALNVSQAWRHSLPYEKVSRLOKACGDAMNLL 357
Qy 357 GYRPVYSEDEQRNLALDLVLPRGLNGFTWASSTASH 392
Db 358 GYRHVSEQEQRNLLLDLL-----STWTVPEQIH 386

Search completed: June 23, 2005, 08:43:27
Job time : 44.6885 secs

THIS PAGE BLANK (USPTO)

GanCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2005, 08:34:49 ; Search time 14.8883 Seconds
(without alignments)
1980.512 Million cell updates/sec

Title: US-10-697-828-13

Perfect score: 2056

Sequence: 1 MWLPRVSTAVTALLAQTF.....LPRGLNGFTWASSTASHPRN 395

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents AA:*
- 1: /cgn2_6/ptodata/1/iaa/5A COMB.pdp.*
 - 2: /cgn2_6/ptodata/1/iaa/5B COMB.pdp.*
 - 3: /cgn2_6/ptodata/1/iaa/6A COMB.pdp.*
 - 4: /cgn2_6/ptodata/1/iaa/6B COMB.pdp.*
 - 5: /cgn2_6/ptodata/1/iaa/PCTUS COMB.pdp.*
 - 6: /cgn2_6/ptodata/1/iaa/backfiles1.pdp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|--|
| 1 | 2056 | 100.0 | 395 | 4 | US-09-949-016-7011 Sequence 7011, Ap |
| 2 | 1729.5 | 84.1 | 390 | 4 | US-09-949-016-6813 Sequence 6813, Ap |
| 3 | 1729.5 | 84.1 | 431 | 4 | US-09-949-016-8893 Sequence 8893, Ap |
| 4 | 1019 | 49.6 | 386 | 3 | US-09-045-284A-2 Sequence 2, Appli |
| 5 | 1019 | 49.6 | 386 | 3 | US-09-190-911-1 Sequence 1, Appli |
| 6 | 981 | 47.7 | 386 | 4 | US-09-786-240-11 Sequence 11, Appl |
| 7 | 663 | 32.2 | 483 | 3 | US-09-263-023-2 Sequence 2, Appli |
| 8 | 663 | 32.2 | 483 | 4 | US-09-471-867-2 Sequence 2, Appli |
| 9 | 657.5 | 32.0 | 608 | 4 | US-09-949-016-9449 Sequence 9449, Ap |
| 10 | 653.5 | 31.8 | 484 | 3 | US-09-263-023-4 Sequence 4, Appli |
| 11 | 653.5 | 31.8 | 484 | 4 | US-09-471-867-4 Sequence 4, Appli |
| 12 | 653.5 | 31.8 | 531 | 4 | US-09-949-016-6471 Sequence 6471, Ap |
| 13 | 576.5 | 28.0 | 479 | 2 | US-08-899-514-2 Sequence 2, Appli |
| 14 | 555.5 | 27.0 | 411 | 3 | US-09-015-188-2 Sequence 2, Appli |
| 15 | 532.5 | 25.9 | 458 | 2 | US-09-655-878-2 Sequence 2, Appli |
| 16 | 100.5 | 4.9 | 199 | 4 | US-09-252-991A-28497 Sequence 28497, A |
| 17 | 96 | 4.7 | 566 | 2 | US-08-484-933B-41 Sequence 41, Appl |
| 18 | 96 | 4.7 | 566 | 2 | US-08-484-158B-41 Sequence 41, Appl |
| 19 | 96 | 4.7 | 566 | 2 | US-08-484-596A-41 Sequence 41, Appl |
| 20 | 96 | 4.7 | 566 | 2 | US-08-480-150A-41 Sequence 41, Appl |
| 21 | 96 | 4.7 | 566 | 3 | US-08-458-731-41 Sequence 41, Appl |
| 22 | 96 | 4.7 | 566 | 3 | US-08-149-223A-41 Sequence 41, Appl |
| 23 | 95.5 | 4.6 | 268 | 4 | US-09-818-780-17 Sequence 17, Appl |
| 24 | 95.5 | 4.6 | 268 | 4 | US-09-818-780-94 Sequence 94, Appl |
| 25 | 93.5 | 4.5 | 575 | 4 | US-09-252-991A-32986 Sequence 32986, A |
| 26 | 93.5 | 4.5 | 1385 | 2 | US-08-687-399-7 Sequence 7, Appli |
| 27 | 93 | 4.5 | 359 | 3 | US-09-150-133-11 Sequence 11, Appl |

ALIGNMENTS

RESULT 1

US-09-949-016-7011
; Sequence 7011, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7011
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-7011

Query Match 100.0%; Score 2056; DB 4; Length 395;

Best Local Similarity 100.0%; Pred. No. 7.5e-210;

Matches 395; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

| | | | |
|----|-----|---|-----|
| Qy | 1 | MWLPVSVSTAVTALLAQTFLLFLVSRPGSPSSPAGGEARVHVLVLSWRS | 60 |
| Db | 1 | MWLPVSVSTAVTALLAQTFLLFLVSRPGSPSSPAGGEARVHVLVLSWRS | 60 |
| Qy | 61 | NQHPDVFYLMPEAHVWVTTLSQCSAATLHMAVRDLVRSVFLCDMDVDA | 120 |
| Db | 61 | NQHPDVFYLMPEAHVWVTTLSQCSAATLHMAVRDLVRSVFLCDMDVDA | 120 |
| Qy | 121 | FQWAVSRALCSPACSAFPRGAISSSEAVCKPLCARQSFLLAREACRSY | 180 |
| Db | 121 | FQWAVSRALCSPACSAFPRGAISSSEAVCKPLCARQSFLLAREACRSY | 180 |
| Qy | 181 | LOVLYPLLSDPALNLRIVHLVRDPRVLRREQTAKALARDNGIVLCTNG | 240 |
| Db | 181 | LOVLYPLLSDPALNLRIVHLVRDPRVLRREQTAKALARDNGIVLCTNG | 240 |
| Qy | 241 | VVREVCESHVRIAEATLKPPPPFLRGYRLVRFPEDLAREPLAEIRALY | 300 |
| Db | 241 | VVREVCESHVRIAEATLKPPPPFLRGYRLVRFPEDLAREPLAEIRALY | 300 |
| Qy | 301 | AWIHNTHGSGPGARREAFKTSRRNALNVSOAWRHALPFAKIRRVQEL | 360 |

| Matches | 216; | Conservative | 45; | Mismatches | 107; | Indels | 28; | Gaps | 7; |
|---------|------|-------------------------|------------------------|-------------------------|------------------|--------|-----|------|----|
| Qy | 14 | LLLAQTFLLLLFLVSRP----- | -GPS | PAGGEARVHVLLVSSWRS | SGSFV | QQL | 59 | | |
| Db | 2 | LLPKMKLLLLFLVSGMAILAF | FHHYSH | NISSLSMKAQPERHVLLVSSWRS | SGSFV | QQL | 61 | | |
| Qy | 60 | FNQHPDVFYLMNEPAHWVWTTLS | QCSAATLHMAVRDLVRSVFLCD | MDVFDAYL-P | WRNLS | 118 | | | |
| Db | 62 | FGQHPDVFYLMNEPAHWVWMTFK | QSTAMWLHMAVRDLIRAVFLCD | MSVFDAYMEF | GP | RRQS | 121 | | |
| Qy | 119 | DLFQWAVSRALCSPPPACSAF | PRGAISSEAVCKPCARQSF | TLAREACRSYSHVVL | KEVRF | 178 | | | |
| Db | 122 | SLFQWENSRALCSAPACDII | IQQDEII | PPRAHCRLLCSQPF | FEVVEKACRSYSHVVL | KEVRF | 181 | | |
| Qy | 179 | FNLOVYPLLLSDPALNLRIVHL | VRDPRAVLRSEQTAKALARD | NGIVLGTGTWV-EADP | 237 | | | | |
| Db | 182 | FNLOSLYPLLKPSLINLHI | VHLVRDPRAVFRSRETKGDL | MIDSRIVMGOHEQKLK | KEDQ | 241 | | | |
| Qy | 238 | GLRVVRECRSHVRIAEAAATLKP | -PPFLRGVRLVRFPEDLARE | PLAEIRALYAF | TGLSLT | 296 | | | |
| Db | 242 | PYYVMQVICSQLEIYK-- | TTQSLPKALOERYLLVRYED | LARAPVAQTSRM | YFVGLEFL | 299 | | | |
| Qy | 297 | PQLEAWIHNIYTHGSGPGARR | EAFKTSNRNALNVSQAMRH | ALPF | AKIRRVQSLCAGAL | 356 | | | |
| Db | 300 | PHLQTVHNIIRGKMGW-- | DHAFHTNAEDALNVSQAW | WLSLYE | KVSRSLQACGD | ANLL | 357 | | |
| Qy | 357 | GYPVYGEDEORNIALDLVL | PRGLNGFTWAS | TASH | 392 | | | | |
| Db | 358 | GYHVRVGEORNILLDLL----- | STWTVPEQIH | 386 | | | | | |

```

RESULT 5
US-09-190-911-1
; Sequence 1, Application US/09190911
; Patent No. 6365365
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Tangemann, Kirsten
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CIP
; CURRENT APPLICATION NUMBER: US/09/190,911
; CURRENT FILING DATE: 1998-11-12
; EARLIER APPLICATION NUMBER: 09/045,284
; EARLIER FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 386
; TYPE: PRT
; ORGANISM: H. sapiens
US-09-190-911-1

Query Match          49.6%; Score 1019; DB 3; Length 386;
Best Local Similarity 54.5%; Pred. No. 1.4e-99;
Matches 216; Conservative 45; Mismatches 107; Indels 28; Gaps 7;

Qy      14  LLLAQTFLLEFLVSRP-----GPSSPAGCEARVHVLVLSWSRSGSFFVQGL 59
Db      2  LLPKQKLLLVLSQMATLAFHMYSHNISLSMKAPQERHVLVLSWSRSGSFFVQGL 61

Qy      60  FNQHPDVPYLMPEPAWHVWTTLSQGSAAATLHMVDRDLVRSVFLCDMDVFDAYL-PWRRNLS 118
Db      62  FQGHDPDVPYLMPEPAWHVWMTFKQSTAMWLHMVDRDLRAVFLCDMSVFDAYMEPGPRQS 121

Qy      119  DLFOWAVSRALCSPPACSAFPRGATSSBAVCKPLCAROSFTLLAREACRSYSHVVLKEVRF 178
Db      122  SLFQWENSRALCSAPACDIIIFQDEIIIPRAKELLCSQPFVEVEKACRSYSHVVLKEVRF 181

Qy      179  FNLQVLYPLLDPPALNLRIVHLVRDPRAVLRSEQTAKALARDNGIVLTGTWV-EADP 237
Db      182  FNLQSLYPLLDKPSLNLRHIVHLVRDPRAVFRSRETKGDLMDISRIVMGOEOKLKKEDQ 241

```

```

Qy 238 GLRVVRVCSHVRIABAATLKP-PPFLRGYRLVRFEDLAREPLAIRALYAFGTGLSLT 296
Db 242 PYTVMQVICOSEIYK--TIQSLPKALQBYLLVRYEDLARAFAVAQTSRMVYFVGLGLEFL 299
Qy 297 POLEAWTHNTTHSGGPCARREAPKTSRRNALNVSQAWRHLPFAKIRRVQELCAGALQLL 356
Db 300 PHLTQVHNITRGKMG--DHAFTNARDALNVSQAWRSLPYEKVSRLOKACGDAMNLL 357
Qy 357 GYRPVYSEDEQRNALDLVLPRGLNGFTWASSTASH 392
Db 358 GYRHRVSEQQRNLLDLL-----STWTVPEQIH 386

RESULT 6
US-09-786-240-11
; Sequence 11, Application US/09786240
; Patent No. 6558935
; GENERAL INFORMATION:
; APPLICANT: INCYTE PHARMACEUTICALS, INC.
; APPLICANT: TANG, Y. Tom
; APPLICANT: CORLEY, Neil C.
; APPLICANT: GUEGLER, Karl J.
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: LAL, Preeti
; APPLICANT: YUS, Henry
; APPLICANT: HILLMAN, Jennifer L.
; APPLICANT: AZIMZAI, Yalda
; TITLE OF INVENTION: HUMAN TRANSFERASE PROTEINS
; FILE REFERENCE: PF-0592 PCT
; CURRENT APPLICATION NUMBER: US/09/786,240
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 09/150,657; unassigned; 60/133,642
; PRIOR FILING DATE: 1998-09-10; 1998-09-10; 1998-11-04; 1999-05-11
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PERL Program
; SEQ ID NO 11
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURES:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6558935 2617407CD1
US-09-786-240-11

```

| Query Match | 47.7% | Score 981; | DB 4; | Length 386; |
|-----------------------|------------------|--|------------|-------------|
| Best Local Similarity | 54.2%; | Pred. No. 1.5e-95; | | |
| Matches 215; | Conservative 42; | Mismatches 110; | Indels 30; | Gaps , 9 |
| QY | 14 | LLLAQTLLPLVSRP-----CPSSPAGGEARVHVLLVLSWSRSGSFFVQQL | 59 | |
| Db | 2 | LLPKOMKLLLVLSQMAILAFPMYSHNTSSLSMKQAQPERMHVLLVLSWSRSGSFFVQQL | 61 | |
| QY | 60 | FNQHPDVFYLMPEAWHWMTTLSQGSAAATLHMVARDLVRSVFLCDMDVFDAYL-PWRRNLS | 118 | |
| Db | 62 | FGQHPDVFYLMPEAWHWMTTFKQSTAMWLHMVARDLIRAVFLCDMSVFDAYMEPPRRQS | 121 | |
| QY | 119 | DLFQWAVSRALCSPPCASAPRPGAISSEAVCKPCLCARQS-FTLAREACRSYSHVVYLKEVR | 177 | |
| Db | 122 | SLFQWENSRALCSPACDII PQDE-SSPGLTAGSCAVNPSPLKLEKACRSYSHVVYLKEVR | 180 | |
| QY | 178 | FFNLQVLYPLTSDPALNLRIVHLVRDPRAVLRSEQTAKALARDNGIVLGNTGTWV-EAD | 236 | |
| Db | 181 | FFNLQVLYPLLKQPSLNLIHVHLVRDPRAVFRSRERTKGDLMIDSRIVMGOHQEKLKED | 240 | |
| QY | 237 | PGLRVVREVCRSYRVIAEAAATLKP-PPLFLKGRYRLVRFEDLAREPLABEIRALYFTGLSL | 295 | |
| Db | 241 | QPYVMQVYCQSQLEIYK--TIQSLPKALQERYLLVRYEDLARAPVAQTSRMIEVFVGLF | 298 | |
| QY | 296 | TPQLEAWIHNI THGSGPGARREAFKTSRNALNYSQAWRHALPFAKTRRVQELCNALQL | 355 | |
| Db | 299 | LPHLOTWVHNITRGKMG--DHAEHTNARDALNYSQAWRSLPYEKVSRLOKACGCDAMNL | 356 | |


```

; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 9449
; LENGTH: 608
; TYPE: PRT
; ORGANISM: Human
; US-09-949-016-9449

Query Match          32.0%; Score 657.5; DB 4; Length 608;
Best Local Similarity 38.2%; Pred. No. 7.6e-61;
Matches 147; Conservative 70; Mismatches 125; Indels 43; Gaps 10;

Qy 30 GPSPAG-----GEARVHVLVLSWSRSGSFVQGLFNQHPDVFYLMPEPAHVVWTTL 80
Db 223 GVAAPCGNTRGTGGVGDKQLVVFTTWSRSGSFVQGLFNQHPDVFYLMPEPAHVVWTTL 282

Qy 81 SQGSAATHMAVRLVRSVFLCDMDVFDAYLP---WRRNLSDL---FQWAVSRALCSPAC 135
Db 283 YPGDAVSLOGAARDMLSALYRCDLSVFLYSPAGSGGRNLTTLGIFGAATNKVVCSSPLC 342

Qy 136 SAPPRGAIS--SEAVCKPLCARQSFLLAREACRSYSHVVLKEVRFNQLVLYPLSDPAL 193
Db 343 PAYRKEVVGVLVDCKKK-CPQQLARFEBCRKYRTLVINGVRVFDVAVLAPLRLDPAL 401

Qy 194 NLRIVHLVRDPRAVLRSGEQTAKALARDN-----GIVLGT--NGT 231
Db 402 DLKVIHLVRDPRAVASRRHSRHLRESLQVVRSDPRAHMPPFLEAGHKLGAKEGV 461

Qy 232 WZEAD-PGLRVVRVCHSHVRIABAAATLKPPFLRGYRLVRFEDLAREPLAEIRALYAF 290
Db 462 GGPADYHALGAMVICNSMAKTLQTA-LQPPDMLOQHYLVVRYEDLVGDPVKTLRRVYDF 520

Qy 291 TGLSLTLQLEAHIWNIHSGPGARRAEAFKTSRNALNVSOAWRHLPFAKIRRVQELCA 350
Db 521 VGLVSPMEQFALNMTSGSSSK--PFVVSARNATQAAWARTALTFOQIKQVEEFCY 578

Qy 351 GALQLIGYRPVYSDEQNALDLV 375
Db 579 QPMAVLGYERVNSPEEVKDLSTLL 603

RESULT 10
US-09-263-023-4
; Sequence 4, Application US/09263023
; Patent No. 6037159
; GENERAL INFORMATION:
; APPLICANT: Uchimura, Kenji
; APPLICANT: Muramatsu, Hideki
; APPLICANT: Kadomatsu, Kenji
; APPLICANT: Kannagi, Reiji
; APPLICANT: Habuchi, Osami
; APPLICANT: Muramatsu, Takashi
; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-SULFOTRANSFERASE AND
; FILE REFERENCE: TOYAMA1.001AUS
; CURRENT APPLICATION NUMBER: US/09/263,023
; EARLIER FILING DATE: 1999-03-05
; EARLIER APPLICATION NUMBER: JP 10-54007
; EARLIER FILING DATE: 1998-03-05
; EARLIER APPLICATION NUMBER: JP 10-177844
; EARLIER FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-263-023-4

Query Match          31.8%; Score 653.5; DB 3; Length 484;
Best Local Similarity 38.8%; Pred. No. 1.4e-60;
Matches 145; Conservative 70; Mismatches 124; Indels 35; Gaps 10;

Qy 33 SPAG-GEARVHVLVLSWSRSGSFVQGLFNQHPDVFYLMPEPAHVVWTTLSQGSATLHMA 91
Db 110 APEGVGDKRHHVMYVFTTWSRSGSFVQGLFNQHPDVFYLMPEPAHVVWTTLSQGSATLHMA 169

Qy 92 VRDLVRSVFLCDMDVFDAYLP---WRRNLSDL---FQWAVSRALCSPACSAFPRGAIS-- 144
Db 170 ARDMLSALYRCDLSVFLYSPAGSGGRNLTTLGIFGAATNKVVCSSPLCPAYRKEVVGVLV 229

Qy 145 SEAVCKPLCARQSFLLAREACRSYSHVVLKEVRFNQLVLYPLSDPALNLRIVHLVRDP 204
Db 230 DDRVCKK-CPQQLARFEBCRKYRTLVINGVRVFDVAVLAPLRLDPALDKVIHLVRDP 288

Qy 205 RAVLRSGEQTAKALARDN-----GIVLGT--NGTWEAD-PGLRV 241
Db 289 RAVSSRIKRSRHLRESLQVVRSDPRAHMPPFLEAGHKLGAKEGVGPGADYHALGA 348

Qy 242 RVSVCHSHVRIABAAATLKPPFLRGYRLVRFEDLAREPLAEIRALYAFGLSLTPOLEA 301
Db 349 MEVICNSMAKTLQTA-LQPPDMLOQHYLVVRYEDLVGDPVKTLRRVYDFVGLLVSPMEQ 407

Qy 302 WIHNIHSGPGARRAEAFKTSRNALNVSOAWRHLPFAKIRRVQELCAGALQLLGYRPV 361
Db 408 FALNMTSGSSSK--PFVVSARNATQAAWARTALTFOQIKQVEEFCYQPMVAVLGYERV 465

Qy 362 YSEDEQNALDLV 375
Db 466 NSPEEVKDLSTLL 479

RESULT 11
US-09-471-867-4
; Sequence 4, Application US/09471867
; Patent No. 6455289
; GENERAL INFORMATION:
; APPLICANT: Uchimura, Kenji
; APPLICANT: Muramatsu, Hideki
; APPLICANT: Kadomatsu, Kenji
; APPLICANT: Kannagi, Reiji
; APPLICANT: Habuchi, Osami
; APPLICANT: Muramatsu, Takashi
; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-SULFOTRANSFERASE AND
; FILE REFERENCE: TOYAMA41.001AUS
; CURRENT APPLICATION NUMBER: US/09/471,867
; CURRENT FILING DATE: 1999-12-23
; EARLIER FILING DATE: 1999-03-05
; EARLIER APPLICATION NUMBER: JP 10-54007
; EARLIER FILING DATE: 1998-03-05
; EARLIER APPLICATION NUMBER: JP 10-177844
; EARLIER FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-471-867-4

Query Match          31.8%; Score 653.5; DB 4; Length 484;
Best Local Similarity 38.8%; Pred. No. 1.4e-60;
Matches 145; Conservative 70; Mismatches 124; Indels 35; Gaps 10;

Qy 33 SPAG-GEARVHVLVLSWSRSGSFVQGLFNQHPDVFYLMPEPAHVVWTTLSQGSATLHMA 91
Db 110 APEGVGDKRHHVMYVFTTWSRSGSFVQGLFNQHPDVFYLMPEPAHVVWTTLSQGSATLHMA 169

Qy 92 VRDLVRSVFLCDMDVFDAYLP---WRRNLSDL---FQWAVSRALCSPACSAFPRGAIS-- 144
Db 170 ARDMLSALYRCDLSVFLYSPAGSGGRNLTTLGIFGAATNKVVCSSPLCPAYRKEVVGVLV 229

Qy 145 SEAVCKPLCARQSFLLAREACRSYSHVVLKEVRFNQLVLYPLSDPALNLRIVHLVRDP 204

```

```
Db 230 DRVCCK-CPQBLARFEECRKYRTLVIKGVRFVDVAVLAPLLRDPALDLKVIHLVRDP 288
QY 205 RAVLRREQTAKALARDN-----GIVLGT--NGTWVAD-PGLRV 241
Db 289 RAVASRIISRHLIRSLQVRSRDPRAHRMPFLEAAGHKLGAKKEGVGGPDADYHALGA 348
QY 242 VREVCRRSHVRIABAATLKPPFFLRGRYLRVRFEDLAREPLAEIRALYAFYGLSLTPOLEA 301
Db 349 MEVICNSMAKTLQTA-LQPPDMLQGHVLYVRVEDLVGDPVKTLRRVYDFVGLLVSPEME 407
QY 302 WIINITHSGSPGARREAFKTSRNALNVSOAQRHALPFAKIRRVQELCAGALQLLGYRPV 361
Db 408 FALNMTSGSGSSK--PFVVSARNATQAANAWRTALTFOQIKQVEEFYQPMVAVLGYRV 465
QY 362 YSEDEQRNLALDLV 375
Db 466 NSPEEVKDLKSKTLL 479

RESULT 12
US-09-949-016-6471
; Sequence 6471, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: GL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 6471
; LENGTH: 531
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6471

Query Match 31.8%; Score 653.5; DB 4; Length 531;
Best Local Similarity 38.8%; Pred. No. 1.6e-60;
Matches 145; Conservative 70; Mismatches 124; Indels 35; Gaps 10;

QY 33 SPAG-GEARVHVLVLSWRSGSFVQGFNFQHPDVFYLMPEPAWHVWTTLSQGSAAITLHWA 91
Db 157 APEGVDKRWYVFTWRSRSGSFVQGFNFQHPDVFYLMPEPAWHVWTTLSQGSAAITLHWA 216
QY 92 VRDLRSVFLCDMDVFDAYLP---WRNLSDL--FQWAVSRALCSPAPCSAFPRAIS-- 144
Db 217 ARDMLSALYRCDLSVFLQSPAGSGGRNLTGLFGAATNKVVCSSPLCPAYRKEVVGVLV 276
QY 145 SEAVCKPLCARQSFYTLAREACRSYHVVLKEVRFNFQVLYPLLSDPALNLRIVHLVRDP 204
Db 277 DRVCCK-CPQBLARFEECRKYRTLVIKGVRFVDVAVLAPLLRDPALDLKVIHLVRDP 335
QY 205 RAVLRREQTAKALARDN-----GIVLGT--NGTWVAD-PGLRV 241
Db 336 RAVASRIISRHLIRSLQVRSRDPRAHRMPFLEAAGHKLGAKKEGVGGPDADYHALGA 395
QY 242 VREVCRRSHVRIABAATLKPPFFLRGRYLRVRFEDLAREPLAEIRALYAFYGLSLTPOLEA 301
Db 396 MEVICNSMAKTLQTA-LQPPDMLQGHVLYVRVEDLVGDPVKTLRRVYDFVGLLVSPEME 454
QY 302 WIINITHSGSPGARREAFKTSRNALNVSOAQRHALPFAKIRRVQELCAGALQLLGYRPV 361
Db 455 FALNMTSGSGSSK--PFVVSARNATQAANAWRTALTFOQIKQVEEFYQPMVAVLGYRV 512
QY 362 YSEDEQRNLALDLV 375
```

```
Db 513 NSPEEVKDLKSKTLL 526

RESULT 13
US-08-899-514-2
; Sequence 2, Application US/08899514
; Patent No. 5910581
; GENERAL INFORMATION:
; APPLICANT: HABUCHI, OSAMI
; APPLICANT: FUKUTA, MASAKAZU
; TITLE OF INVENTION: POLYPEPTIDE OF GLYCOSAMINOGLYCAN
; TITLE OF INVENTION: SULFOTRANSFERASE ORIGINATING FROM HUMAN AND DNA CODING
; TITLE OF INVENTION: FOR THE POLYPEPTIDE
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: KNOBEE, MARTENS, OLSON & BEAR, LLP
; STREET: 620 NEWPORT CENTER DRIVE, SIXTEENTH FLOOR
; CITY: NEWPORT BEACH
; STATE: CALIFORNIA
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/899,514
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: DANIEL E ALTMAN
; REGISTRATION NUMBER: 34,115
; REFERENCE/DOCKET NUMBER: TOYAM21.001AUS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 714 760 0404
; TELEFAX: 714 760 9502
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 479
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-899-514-2

Query Match 28.0%; Score 576.5; DB 2; Length 479;
Best Local Similarity 37.2%; Pred. No. 2.2e-52;
Matches 136; Conservative 58; Mismatches 141; Indels 31; Gaps 8;

QY 31 PPSPPAGEARVHVLVLSWRSGSFVQGFNFQHPDVFYLMPEPAWHVWTTLS-----QGSAA 86
Db 123 PRPFAVAGPRRHVLLMATRTGSSFVGFENQGNIFYLFEPLWHIERTVTSFPPGGANAA 182
QY 87 TLHMAVRDLVRSVFLCDMDVFDAY---LPWRRNLSDLFQWAVSRALCSPAPCSAFPRAI 143
Db 183 GSALVYRDVLKQLFLCDLYVLEHFIPLPEDHLTQPMFRGSSRSLCEDPVCTPFVKKVF 242
QY 144 SSEAVCKPLCARQSFYTLAREACRSYHVVLKEVRFNFQVLYPLLSDPALNLRIVHLVRD 203
Db 243 EKHCKNRKRGPLUNVTAAEACRKEHMAKAVIRIQLEFLQPLAEDPRLDLRVQLVRD 302
QY 204 PRAVLRREQTAKALARDNGIVLGTNGTW-----EADPGLRV-----VREVCRRSHVRIA 253
Db 303 PRAVLASR---MVAPA-----GKYTKWKWLDDEGQDGLREEVQRLRGNCES-IRLS 351
QY 254 EAATLKPPFPFLRGYRLVRFEDLAREPLAEIRALYAFYGLSLTPOLEAHTHNTGSGGPG 313
Db 352 AELGLRQPAWLGRYMLVRYEDVARGFLQKAREMYPFAGIPLTPQVEDWTQKNTQAAHDG 411
```

Qy 314 ARREAFKTSRNALNVSAWRHALPFKIRRVQELCAGALQLLGYRPVYSEQRNLALD 373
Db 412 S--GIYSTQKNSSEKQKRFSPFKLAQVQAPCGPAMRLFGYKLARDAALNRSVS 468
Qy 374 LVLPRG 379
Db 469 LLEERG 474

RESULT 14

US-09-015-188-2
; Sequence 2, Application US/09015188C
; Patent No. 6399358
; GENERAL INFORMATION:
; APPLICANT: Williams, Kevin J
; APPLICANT: Tabas, Ira
; TITLE OF INVENTION: A Human Gene Encoding Human Chondroitin
; FILE OF INVENTION: 6-Sulfotransferase
; FILE REFERENCE: JEFF-0231
; CURRENT APPLICATION NUMBER: US/09/015.188C
; CURRENT FILING DATE: 1998-01-29
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 2
; LENGTH: 411
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-015-188-2

Query Match 27.0%; Score 555.5; DB 3; Length 411;
Best Local Similarity 36.7%; Pred. No. 2.9e-50;
Matches 131; Conservative 56; Mismatches 153; Indels 17; Gaps 8;
Qy 40 RVHVLSSWRSGSFVQLENQHPDVFYLMPEAHVWTTL-----ATLHMA 91
Db 60 KTHILILATTSRSGSFVQLENQHLDFYLFPELYHVQNTLIPRTQKSPADRRVMLGA 119
Qy 92 VRDLVRSVFLCDMDVFDAYL---PWRNLSDLFQWAVSRALCSPACS-APPRCAISSEA 147
Db 120 SRDLRSYDCLYDFLENYIKPPVNHHTDIFRGRASVLCSPVCDPPGADLVLEEG 179
Qy 148 VCKPLCARQSTFLAREACRSYSHVLYKEVRFNQLVLYPLSDPALNLRVHLVRDPRV 207
Db 180 DCVRKCGLLNLTVAAEACRSHVAIKTVRVPEVNDLRALVEDPRLNLKVQLVRDPGI 239
Qy 208 LRSREQTAKALARDNGIVLGTNGTWREADPGLRVVREVCRSVHRIAENATLKPPLRGR 267
Db 240 LASRSETPRDYRLWRLWYGTGRKPYNLD--VTQLTTVCEDFSNSVSTGLMR--PPWLKKG 296
Qy 268 YRLVRFEDLAREPLAEIRALYAFGLSLTPOLEAWIHNITHGSGPGARREAFKTSRNAL 327
Db 297 YMLVRYEDLANPMKKTETIYGFGLIPDLSHVARIWQNTTRGD--PTLGKHKYGTVRNSA 354
Qy 328 NVSAWRHALPFKIRRVQELCAGALQLLGYRPVYSEQRNLALDLVLPRLNGFT 384
Db 355 ATAENRPLSYDIVAFQAQACQVLAQLGYKIAASEELKNPSVLSVEERDFRPS 411

RESULT 15

US-08-655-878-2
; Sequence 2, Application US/08655878
; Patent No. 582713
; GENERAL INFORMATION:
; APPLICANT: FUKUTA, MASAKAZU
; APPLICANT: HABUCHI, OSAMI
; TITLE OF INVENTION: DNA CODING FOR SULFOTRANSFERASE
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE:
; STREET:
; CITY:
; STATE:
; COUNTRY:

; ZIP:
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/655,878
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME:
; REGISTRATION NUMBER:
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE:
; TELEFAX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 458
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-655-878-2

Query Match 25.9%; Score 532.5; DB 2; Length 458;
Best Local Similarity 34.8%; Pred. No. 9.7e-48;
Matches 134; Conservative 63; Mismatches 157; Indels 33; Gaps 10;
Qy 5 RVSTTAVTALLAQTFLLFLVSRPGPSSPAGCEARVHVLVLSWSRSGSFVQGFQHP 64
Db 80 RQDTHAFSQRRLRNVTQLAGELIAP---EPRHVLMMATRTGSSFVGEFFNQOQ 136
Qy 65 DVFYLMPEAHV---WTTLSQGSAAITLHMAV---BDLVRSVFLCDMDVFDAYL---PWRNL 117
Db 137 NIFLFEPLWHIERTVTFEPGGANAVGSALVYRDVLQQLLCLDYILESFISPAPEHLT 196
Qy 118 SDFQWAVSRALCSPACSAFPRAISSEAVCKPLCARQSTFLAREACRSYSHVLYKEVR 177
Db 197 AALFRGSSHSLCEEPVCTPSLKKVFEKYHCKNRCGCLNITLAAEACRRKQHWALKTVR 256
Qy 178 FFNLQVLYPLSDPALNLRVHLVRDPRVLRVRSREQTAKALARDNGIVLGTNGTW---V 233
Db 257 IROLEFLQPLAEDPRLDLRIQLVRDPRVLRVSRWAFS-----GKYESWKKWAA 306
Qy 234 EADPGRLV-----VREVCRSVHRIAENATLKPPLFRLGRYLRVRFEDLAREPLAEIRALY 288
Db 307 EGEAPLQDEDEVQRLGNCS-IRLSAELGLRQPRWLRGRYMLVRYEDVARAPLRKALEMY 365
Qy 289 AFTGLSLTPOLEAWIHNITHGSGPGARREAFKTSRNALNVSAWRHALPFKIRRVQEL 348
Db 366 RFAGIHPTQVEEWIRANT--QAQDSNGIYST--QKNSSEQFEKWFESIPPKLAQVQDA 422
Qy 349 CAGALQLLGYRPVYSEQRNLALDLV 375
Db 423 CEPAMRLFGYKLASSAQELNRSLSLL 449

Search completed: June 23, 2005, 08:52:14
Job time : 15.8893 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2005, 08:49:14 ; Search time 49.935 Seconds
(without alignments)
3041.886 Million cell updates/sec

Title: US-10-697-828-13

Perfect score: 2056

Sequence: 1 MWLPRVSTAVTALLAQTF.....LPRGLNGFTWASSTASHPRN 395

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1717557 seqs, 384547976 residues

Total number of hits satisfying chosen parameters: 1717557

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications AA:

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US10A_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/1/pubpaa/US11A_PUBCOMB.pep.*
- 20: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
- 21: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 22: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|---------------------|
| 1 | 2056 | 100.0 | 395 | 9 | US-09-927-602-2 |
| 2 | 2056 | 100.0 | 395 | 15 | US-10-258-080-5 |
| 3 | 2056 | 100.0 | 395 | 16 | US-10-648-593-159 |
| 4 | 2056 | 100.0 | 395 | 16 | US-10-697-828-13 |
| 5 | 1934 | 94.1 | 395 | 9 | US-09-927-602-3 |
| 6 | 1729.5 | 84.1 | 390 | 9 | US-09-927-602-4 |
| 7 | 1729.5 | 84.1 | 390 | 16 | US-10-697-828-8 |
| 8 | 1708 | 83.1 | 395 | 16 | US-10-697-828-7 |
| 9 | 1708 | 83.1 | 395 | 16 | US-10-841-707-8 |
| 10 | 1694.5 | 82.4 | 418 | 9 | US-09-927-602-5 |
| 11 | 1538 | 74.8 | 394 | 16 | US-10-408-765A-2211 |

| | | | | | | |
|----|-------|------|------|----|----------------------|--------------------|
| 12 | 1019 | 49.6 | 386 | 9 | US-09-916-825-2 | Sequence 2, Appli |
| 13 | 1019 | 49.6 | 386 | 13 | US-10-007-262-1 | Sequence 1, Appli |
| 14 | 1018 | 49.5 | 380 | 16 | US-10-841-707-6 | Sequence 6, Appli |
| 15 | 981 | 47.7 | 386 | 14 | US-10-427-631-11 | Sequence 11, Appli |
| 16 | 808 | 39.3 | 169 | 9 | US-09-927-602-6 | Sequence 6, Appli |
| 17 | 755 | 36.7 | 189 | 9 | US-09-927-602-7 | Sequence 7, Appli |
| 18 | 663 | 32.2 | 483 | 14 | US-10-212-933-2 | Sequence 2, Appli |
| 19 | 657.5 | 32.0 | 530 | 16 | US-10-723-860-1409 | Sequence 1409, Ap |
| 20 | 653.5 | 31.8 | 484 | 14 | US-10-212-933-4 | Sequence 4, Appli |
| 21 | 653.5 | 31.8 | 531 | 9 | US-09-933-790-255 | Sequence 255, App |
| 22 | 653.5 | 31.8 | 531 | 16 | US-10-755-889-284 | Sequence 284, App |
| 23 | 578.5 | 28.1 | 479 | 13 | US-10-087-192-126 | Sequence 126, App |
| 24 | 569 | 27.7 | 481 | 13 | US-10-087-192-128 | Sequence 123, App |
| 25 | 555.5 | 27.0 | 411 | 14 | US-10-021-660-128 | Sequence 128, App |
| 26 | 555.5 | 27.0 | 411 | 15 | US-10-211-462-97 | Sequence 97, Appli |
| 27 | 555.5 | 27.0 | 411 | 16 | US-10-408-765A-395 | Sequence 395, App |
| 28 | 555.5 | 27.0 | 411 | 16 | US-10-723-860-1544 | Sequence 1544, Ap |
| 29 | 521.5 | 25.4 | 171 | 9 | US-09-927-602-8 | Sequence 8, Appli |
| 30 | 377 | 18.3 | 169 | 9 | US-09-927-602-9 | Sequence 9, Appli |
| 31 | 364.5 | 17.7 | 174 | 9 | US-09-927-602-11 | Sequence 11, Appli |
| 32 | 348 | 16.9 | 179 | 9 | US-09-927-602-10 | Sequence 10, Appli |
| 33 | 328 | 16.2 | 387 | 14 | US-10-126-279-21 | Sequence 21, Appli |
| 34 | 328 | 16.2 | 387 | 14 | US-10-286-606-21 | Sequence 21, Appli |
| 35 | 328 | 16.2 | 387 | 16 | US-10-891-383-21 | Sequence 21, Appli |
| 36 | 123.5 | 6.0 | 596 | 16 | US-10-697-828-9 | Sequence 9, Appli |
| 37 | 123.5 | 6.0 | 1212 | 17 | US-10-479-472A-2 | Sequence 2, Appli |
| 38 | 123.5 | 6.0 | 1222 | 16 | US-10-697-828-15 | Sequence 15, Appli |
| 39 | 123.5 | 6.0 | 1223 | 16 | US-10-475-446-4 | Sequence 4, Appli |
| 40 | 118.5 | 5.8 | 1207 | 16 | US-10-697-828-17 | Sequence 17, Appli |
| 41 | 118.5 | 5.8 | 1207 | 17 | US-10-479-472A-4 | Sequence 4, Appli |
| 42 | 108 | 5.3 | 2588 | 16 | US-10-437-963-138279 | Sequence 138279, A |
| 43 | 107 | 5.2 | 299 | 14 | US-10-156-761-9846 | Sequence 9846, Ap |
| 44 | 98 | 4.8 | 276 | 15 | US-10-425-114-47175 | Sequence 47175, A |
| 45 | 98 | 4.8 | 281 | 16 | US-10-425-115-213557 | Sequence 213557, A |

ALIGNMENTS

RESULT 1
US-09-927-602-2
; Sequence 2, Application US/09927602
; Patent No. US20020061562A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; APPLICANT: Akama, Tomoya O.
; TITLE OF INVENTION: Methods of Treating Macular Corneal
; FILE REFERENCE: P-LJ 4852
; CURRENT APPLICATION NUMBER: US/09/927,602
; CURRENT FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 09/638,211
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-927-602-2

| | | | | |
|-----------------------|-----------------|---|-----------|-------------|
| Query Match | 100.0% | Score 2056; | DB 9; | Length 395; |
| Best Local Similarity | 100.0%; | Pred. No. 4.6e-198; | | |
| Matches 395; | Conservative 0; | Mismatches 0; | Indels 0; | Gaps 0; |
| Qy | 1 | MWLPVSTAVTALLAQTFLLFLVSRPGSPAGGEARVHVLVLSWRSQSGSFVQOLF | 60 | |
| Db | 1 | MWLPVSTAVTALLAQTFLLFLVSRPGSPAGGEARVHVLVLSWRSQSGSFVQOLF | 60 | |
| Qy | 61 | NQHPDVFYLMPEAWHVVTTLSQCSAATLHMVDRVLSVFLCDMDVFDAYLPWRNLSDL | 120 | |
| Db | 61 | NQHPDVFYLMPEAWHVVTTLSQCSAATLHMVDRVLSVFLCDMDVFDAYLPWRNLSDL | 120 | |


```
Db 181 LQVLYPLSDPALNLRIVHLVRDPAVLRSEQTAKALARDNGIVLGTNGTWVADPGLR 240
Qy 241 VVREVCRRSHVRIAEAAATLKPPFLRGYRLVRFPEDLAREPLAEIRALYAFGLSLTPGLE 300
Db 241 VVREVCRRSHVRIAEAAATLKPPFLRGYRLVRFPEDLAREPLAEIRALYAFGLSLTPGLE 300
Qy 301 AWIHNITHSGPGARREAFKTSRRNALNVQAWRHLPFAKIRRVQELCAGALQLLGYRP 360
Db 301 AWIHNITHSGPGARREAFKTSRRNALNVQAWRHLPFAKIRRVQELCAGALQLLGYRP 360
Qy 361 VYSEDEQRNALDLVLPRLNGFTWASSTASHPRN 395
Db 361 VYSEDEQRNALDLVLPRLNGFTWASSTASHPRN 395

RESULT 4
US-10-697-828-13
; Sequence 13, Application US/10697828
; Publication No. US20040185546A1
; GENERAL INFORMATION:
; APPLICANT: Rosen, Steven
; APPLICANT: Lee, Jin Kyu
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: Novel Glycosyl Sulfotransferases GST-4alpha, & GST-6
; FILE REFERENCE: UCAL-138DIV
; CURRENT APPLICATION NUMBER: US/10/697,828
; PRIOR FILING DATE: 2003-10-29
; PRIOR APPLICATION NUMBER: 09/593,828
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: 60/144,694
; PRIOR FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 395
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-697-828-13

Query Match 100.0%; Score 2056; DB 16; Length 395;
Best Local Similarity 100.0%; Pred. No. 4.6e-198;
Matches 395; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MWLPRVSTAVTALLAQTFLLFLVSRPGSPSPAGGEARVHVLVLSWRSWSSGSSFGQLF 60
Db 1 MWLPRVSTAVTALLAQTFLLFLVSRPGSPSPAGGEARVHVLVLSWRSWSSGSSFGQLF 60
Qy 61 NQHPDVFYLMPEPAWHVMTTLLSQGSAATLHMAVRDLVRSVFLCDMDVFDAYLPWRNLSDL 120
Db 61 NQHPDVFYLMPEPAWHVMTTLLSQGSAATLHMAVRDLVRSVFLCDMDVFDAYLPWRNLSDL 120
Qy 121 FQAVSRALCSPACSAFPRAISSEAVCKPLCARQSFLLAREACRSYSHVVLKEVRFN 180
Db 121 FQAVSRALCSPACSAFPRAISSEAVCKPLCARQSFLLAREACRSYSHVVLKEVRFN 180
Qy 181 LQVLYPLSDPALNLRIVHLVRDPAVLRSEQTAKALARDNGIVLGTNGTWVADPGLR 240
Db 181 LQVLYPLSDPALNLRIVHLVRDPAVLRSEQTAKALARDNGIVLGTNGTWVADPGLR 240
Qy 241 VVREVCRRSHVRIAEAAATLKPPFLRGYRLVRFPEDLAREPLAEIRALYAFGLSLTPGLE 300
Db 241 VVREVCRRSHVRIAEAAATLKPPFLRGYRLVRFPEDLAREPLAEIRALYAFGLSLTPGLE 300
Qy 301 AWIHNITHSGPGARREAFKTSRRNALNVQAWRHLPFAKIRRVQELCAGALQLLGYRP 360
Db 301 AWIHNITHSGPGARREAFKTSRRNALNVQAWRHLPFAKIRRVQELCAGALQLLGYRP 360
Qy 361 VYSEDEQRNALDLVLPRLNGFTWASSTASHPRN 395
Db 361 VYSEDEQRNALDLVLPRLNGFTWASSTASHPRN 395

RESULT 5
```

```
US-09-927-602-3
; Sequence 3, Application US/09927602
; Patent No. US20020061562A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; APPLICANT: Akama, Tomoya O.
; TITLE OF INVENTION: Methods of Treating Macular Corneal
; FILE REFERENCE: P-LJ 4852
; CURRENT APPLICATION NUMBER: US/09/927,602
; CURRENT FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 09/638,211
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: synthetic construct
; NAME/KEY: VARIANT
; LOCATION: (1)...(395)
; OTHER INFORMATION: Xaa = any amino acid
US-09-927-602-3

Query Match 94.1%; Score 1934; DB 9; Length 395;
Best Local Similarity 94.4%; Pred. No. 9.1e-186;
Matches 371; Conservative 3; Mismatches 19; Indels 0; Gaps 0;

Qy 1 MWLPRVSTAVTALLAQTFLLFLVSRPGSPSPAGGEARVHVLVLSWRSWSSGSSFGQLF 60
Db 1 MWLPRVSTAVTALLAQTFLLFLVSRPGSPSPAGGEARVHVLVLSWRSWSSGSSFGQLF 60
Qy 61 NQHPDVFYLMPEPAWHVMTTLLSQGSAATLHMAVRDLVRSVFLCDMDVFDAYLPWRNLSDL 120
Db 61 NQHPDVFYLMPEPAWHVMTTLLSQGSAATLHMAVRDLVRSVFLCDMDVFDAYLPWRNLSDL 120
Qy 121 FQAVSRALCSPACSAFPRAISSEAVCKPLCARQSFLLAREACRSYSHVVLKEVRFN 180
Db 121 FQAVSRALCSPACSAFPRAISSEAVCKPLCARQSFLLAREACRSYSHVVLKEVRFN 180
Qy 181 LQVLYPLSDPALNLRIVHLVRDPAVLRSEQTAKALARDNGIVLGTNGTWVADPGLR 240
Db 181 LQVLYPLSDPALNLRIVHLVRDPAVLRSEQTAKALARDNGIVLGTNGTWVADPGLR 240
Qy 241 VVREVCRRSHVRIAEAAATLKPPFLRGYRLVRFPEDLAREPLAEIRALYAFGLSLTPGLE 300
Db 241 VVREVCRRSHVRIAEAAATLKPPFLRGYRLVRFPEDLAREPLAEIRALYAFGLSLTPGLE 300
Qy 301 AWIHNITHSGPGARREAFKTSRRNALNVQAWRHLPFAKIRRVQELCAGALQLLGYRP 360
Db 301 AWIHNITHSGPGARREAFKTSRRNALNVQAWRHLPFAKIRRVQELCAGALQLLGYRP 360
Qy 361 VYSEDEQRNALDLVLPRLNGFTWASSTASHPRN 393
Db 361 VYSEDEQRNALDLVLPRLNGFTWASSTASHPRN 393

RESULT 6
US-09-927-602-4
; Sequence 4, Application US/09927602
; Patent No. US20020061562A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; APPLICANT: Akama, Tomoya O.
; TITLE OF INVENTION: Methods of Treating Macular Corneal
; FILE REFERENCE: P-LJ 4852
; CURRENT APPLICATION NUMBER: US/09/927,602
; CURRENT FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 09/638,211
; PRIOR FILING DATE: 2000-08-11
```

; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 390
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-927-602-4

Query Match 84.1%; Score 1729.5; DB 9; Length 390;
Best Local Similarity 85.8%; Pred. No. 3.6e-165;
Matches 333; Conservative 17; Mismatches 37; Indels 1; Gaps 1;

Qy 1 MWLPRVSVSTAVTALLAQ-TFLLLFLVSRPGSPAGGEARVHVLSSWRSGSSFVGL 59
Db 1 MWLPRFSKTVTVLLLAQTCLLLFIISRPQSPAGGEDRVHVLSSWRSGSSFVGL 60
Qy 60 FNOHPDVFLMEPAWHVWTTLSQGSAAATLHMAVRDLVRSVFLCDMDVDFDAYLWPRNLS 119
Db 61 FSOHPDVFLMEPAWHVWTTLSQGSAAATLHMAVRDLVRSVFLCDMDVDFDAYMPSRNL 120
Qy 120 LFQAVSRALCSPACSAFPRGTSISKQDVCKTLCRTQPFSLAREACRSYSHVVLKEVR 179
Db 121 FFWATSRALCSPACSAFPRGTSISKQDVCKTLCRTQPFSLAREACRSYSHVVLKEVR 180
Qy 180 NLQVLYPLSDPALNLRIVHLVRDPRAVLRSREQTAKALARDNGIVLGTNGTWVEADPGL 239
Db 181 NLQVLYPLSDPALNLRIVHLVRDPRAVLRSREQTAKALARDNGIVLGTNGTWVEADPH 240
Qy 240 RVREVCVSHVRIAEATLKPPPFLLGRYRLVRFEDLAREPLAEIRALYAFTGLSLTPOL 299
Db 241 RLIREVCVSHVRIAEATLKPPPFLLGRYRLVRFEDLAREPLAEIRALYAFTGLSLTPOL 300
Qy 300 EAWIHNTHSGGPGARREAFKTSRNALNVSOAQRHALPFAKIRRVQELCAGALQLLYR 359
Db 301 EAWIHNTHSGGPGARREAFKTSRNALNVSOAQRHALPFAKIRRVQELCAGALQLLYR 360
Qy 360 PVYSEDEQORNALDLVLRGLNGFTWAS 387
Db 361 PVYSADQQRDLTLVLPRGPDHFSWAS 388

RESULT 7

US-10-697-828-8
; Sequence 8, Application US/10697828
; Publication No. US20040185546A1
; GENERAL INFORMATION:
; APPLICANT: Rosen, Steven
; APPLICANT: Lee, Jin Kyu
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: Novel Glycosyl Sulfotransferases GST-4alpha, GST-4beta, & GST-6
; FILE REFERENCE: UCAL-138DIV
; CURRENT APPLICATION NUMBER: US/10/697,828
; CURRENT FILING DATE: 2003-10-29
; PRIOR APPLICATION NUMBER: 09/593,828
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: 60/144,694
; PRIOR FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 390
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-697-828-8

Query Match 84.1%; Score 1729.5; DB 16; Length 390;
Best Local Similarity 85.8%; Pred. No. 3.6e-165;
Matches 333; Conservative 17; Mismatches 37; Indels 1; Gaps 1;

Qy 1 MWLPRVSVSTAVTALLAQ-TFLLLFLVSRPGSPAGGEARVHVLSSWRSGSSFVGL 59
Db 1 MWLPRFSKTVTVLLLAQTCLLLFIISRPQSPAGGEDRVHVLSSWRSGSSFVGL 60

Qy 60 FNOHPDVFLMEPAWHVWTTLSQGSAAATLHMAVRDLVRSVFLCDMDVDFDAYLWPRNLS 119
Db 61 FSOHPDVFLMEPAWHVWTTLSQGSAAATLHMAVRDLVRSVFLCDMDVDFDAYMPSRNL 120
Qy 120 LFQAVSRALCSPACSAFPRGTSISKQDVCKTLCRTQPFSLAREACRSYSHVVLKEVR 179
Db 121 FFWATSRALCSPACSAFPRGTSISKQDVCKTLCRTQPFSLAREACRSYSHVVLKEVR 180
Qy 180 NLQVLYPLSDPALNLRIVHLVRDPRAVLRSREQTAKALARDNGIVLGTNGTWVEADPGL 239
Db 181 NLQVLYPLSDPALNLRIVHLVRDPRAVLRSREQTAKALARDNGIVLGTNGTWVEADPH 240
Qy 240 RVREVCVSHVRIAEATLKPPPFLLGRYRLVRFEDLAREPLAEIRALYAFTGLSLTPOL 299
Db 241 RLIREVCVSHVRIAEATLKPPPFLLGRYRLVRFEDLAREPLAEIRALYAFTGLSLTPOL 300
Qy 300 EAWIHNTHSGGPGARREAFKTSRNALNVSOAQRHALPFAKIRRVQELCAGALQLLYR 359
Db 301 EAWIHNTHSGGPGARREAFKTSRNALNVSOAQRHALPFAKIRRVQELCAGALQLLYR 360
Qy 360 PVYSEDEQORNALDLVLRGLNGFTWAS 387
Db 361 PVYSADQQRDLTLVLPRGPDHFSWAS 388

RESULT 8

US-10-697-828-7
; Sequence 7, Application US/10697828
; Publication No. US20040185546A1
; GENERAL INFORMATION:
; APPLICANT: Rosen, Steven
; APPLICANT: Lee, Jin Kyu
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: Novel Glycosyl Sulfotransferases GST-4alpha, GST-4beta, & GST-6
; FILE REFERENCE: UCAL-138DIV
; CURRENT APPLICATION NUMBER: US/10/697,828
; CURRENT FILING DATE: 2003-10-29
; PRIOR APPLICATION NUMBER: 09/593,828
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: 60/144,694
; PRIOR FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 395
; TYPE: PRT
; ORGANISM: mus musculus
US-10-697-828-7

Query Match 83.1%; Score 1708; DB 16; Length 395;
Best Local Similarity 83.0%; Pred. No. 5.4e-163;
Matches 328; Conservative 24; Mismatches 43; Indels 0; Gaps 0;

Qy 1 MWLPRVSVSTAVTALLAQ-TFLLLFLVSRPGSPAGGEARVHVLSSWRSGSSFVGL 60
Db 1 MWLPRFSKTVTVLLLAQTCLLLFIISRPQSPAGGEDRVHVLSSWRSGSSFVGL 60
Qy 61 FNOHPDVFLMEPAWHVWTTLSQGSAAATLHMAVRDLVRSVFLCDMDVDFDAYLWPRNLS 120
Db 61 FSOHPDVFLMEPAWHVWTTLSQGSAAATLHMAVRDLVRSVFLCDMDVDFDAYLWPRNLS 120
Qy 121 FQAVSRALCSPACSAFPRGTSISKQDVCKTLCRTQPFSLAREACRSYSHVVLKEVR 180
Db 121 FQAVSRALCSPACSAFPRGTSISKQDVCKTLCRTQPFSLAREACRSYSHVVLKEVR 180
Qy 181 NLQVLYPLSDPALNLRIVHLVRDPRAVLRSREQTAKALARDNGIVLGTNGTWVEADPGL 240
Db 181 NLQVLYPLSDPALNLRIVHLVRDPRAVLRSREQTAKALARDNGIVLGTNGTWVEADPH 240
Qy 241 RVREVCVSHVRIAEATLKPPPFLLGRYRLVRFEDLAREPLAEIRALYAFTGLSLTPOL 300
Db 241 VVNEVCVSHVRIAEATLKPPPFLLGRYRLVRFEDLAREPLAEIRALYAFTGLSLTPOL 300

```

Qy 301 AWIHNTHGSGPGARREAFKTSRNALNVSOAWRHLPFAKIRRVQELCAGALQLLYRP 360
Db 301 TWIHNTHGSGPGARREAFKTSRDALSVQAWRHLPFAKIRRVQELCAGALQLLYRS 360
Qy 361 VYSEDEQNLALDLVLPRLNGFTWASSTASHPRN 395
Db 361 VHSLEQDLSLDLLPRGMDSFKWASSTEKQPS 395

RESULT 9
US-10-841-707-8
; Sequence 8, Application US/10841707
; Publication No. US20040202649A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Minoru
; APPLICANT: Yeh, Jiunn-Chern
; APPLICANT: Hiraoka, Nobuyoshi
; TITLE OF INVENTION: Identification of the Meca-79 Antigen
; TITLE OF INVENTION: and Related Methods of Treating L-Selectin-Mediated
; TITLE OF INVENTION: Conditions
; FILE REFERENCE: P-LJ 4149
; CURRENT APPLICATION NUMBER: US/10/841,707
; CURRENT FILING DATE: 2004-05-05
; PRIOR APPLICATION NUMBER: US/09/569,320A
; PRIOR FILING DATE: 2000-05-11
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-841-707-8

Query Match 83.1%; Score 1708; DB 16; Length 395;
Best Local Similarity 83.0%; Pred. No. 5.4e-163;
Matches 328; Conservative 24; Mismatches 43; Indels 0; Gaps 0;

Qy 1 MWLPRVSSTAVTALLAQTFLLFLVSRPGSPAGGEARVHVLVLSWRSWGSSFGVQLF 60
Db 1 MRLPRFSSTVMSLLMVQTGILVLSRQVPSPAGLGERVHVLVLSWRSWGSSFGVQLF 60
Qy 61 NQHPDVFLMEPAWHVMTTLSQGSAAHLHMAVRDLVRSVFLCDMDVFDAYLPWRNLSDL 120
Db 61 SQHPDVFLMEPAWHVMTTLSQGSAPALHMAVRDLIRSVFLCDMDVFDAYLPWRNLSDL 120
Qy 121 FQWAVSRALCSPPCSAFPAGISSEAVCKPLCARQSFLLAREACRSYSHVVLKEVRFN 180
Db 121 FQWAVSRALCSPPCSAFPAGISSEAVCKPLCATRPFGLAQACSSYSHVVLKEVRFN 180
Qy 181 LQVLYPLSDPALNLRIVHLVRDPRAVLRSEQTAKALARDNGIVLGTNGTWVEADPGLR 240
Db 181 LQVLYPLSDPALNLRIVHLVRDPRAVLRSEQTAKALARDNGIVLGTNGTWVEADPGLR 240
Qy 241 VVREYCRSHVRIABEATLKPPPLRGYRLVRFDLAREPLAEIRALYAFGLSLTPQLE 300
Db 241 VVNEVCRSVRIABEALHKPPPLQDRYRLVRYEDLARDPLTVIRELYAFGLSLTPQLQ 300
Qy 301 AWIHNTHGSGPGARREAFKTSRNALNVSOAWRHLPFAKIRRVQELCAGALQLLYRP 360
Db 301 TWIHNTHGSGPGARREAFKTSRDALSVQAWRHLPFAKIRRVQELCAGALQLLYRS 360
Qy 361 VYSEDEQNLALDLVLPRLNGFTWASSTASHPRN 395
Db 361 VHSLEQDLSLDLLPRGMDSFKWASSTEKQPS 395

RESULT 10
US-09-927-602-5
; Sequence 5, Application US/09927602
; Patent No. US20020061562A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; APPLICANT: Akama, Tomoya O.

```

```

; TITLE OF INVENTION: Methods of Treating Macular Corneal
; TITLE OF INVENTION: Dystrophy
; FILE REFERENCE: P-LJ 4852
; CURRENT APPLICATION NUMBER: US/09/927,602
; CURRENT FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 09/638,211
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 418
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-927-602-5

Query Match 82.4%; Score 1694.5; DB 9; Length 418;
Best Local Similarity 82.8%; Pred. No. 1.3e-161;
Matches 327; Conservative 24; Mismatches 43; Indels 1; Gaps 1;

Qy 1 MWLPRVSSTAVTALLAQTFLLFLVSRPGSPAGGEARVHVLVLSWRSWGSSFGVQLF 60
Db 25 MRLPRFSSTVMSLLMVQTGILVLSRQVPSPAGLGERVHVLVLSWRSWGSSFGVQLF 84
Qy 61 NQHPDVFLMEPAWHVMTTLSQGSAAHLHMAVRDLVRSVFLCDMDVFDAYLPWRNLSDL 120
Db 85 SQHPDVFLMEPAWHVMTTLSQGSAPALHMAVRDLIRSVFLCDMDVFDAYLPWRNLSDL 144
Qy 121 FQWAVSRALCSPPCSAFPAGISSEAVCKPLCARQSFLLAREACRSYSHVVLKEVRFN 180
Db 145 FQWAVSRALCSPPCSAFPAGISSEAVCKPLCATRPFGLAQACSSYSHVVLKEVRFN 204
Qy 181 LQVLYPLSDPALNLRIVHLVRDPRAVLRSEQTAKALARDNGIVLGTNGTWVEADPGLR 240
Db 205 LQVLYPLSDPALNLRIVHLVRDPRAVLRSEQTAKALARDNGIVLGTNGTWVEADPGLR 264
Qy 241 VVREYCRSHVRIABEATLKPPPLRGYRLVRFDLAREPLAEIRALYAFGLSLTPQLE 300
Db 265 VVNEVCRSVRIABEALHKPPPLQDRYRLVRYEDLARDPLTVIRELYAFGLSLTPQLQ 323
Qy 301 AWIHNTHGSGPGARREAFKTSRNALNVSOAWRHLPFAKIRRVQELCAGALQLLYRP 360
Db 324 TWIHNTHGSGPGARREAFKTSRDALSVQAWRHLPFAKIRRVQELCAGALQLLYRS 383
Qy 361 VYSEDEQNLALDLVLPRLNGFTWASSTASHPRN 395
Db 384 VHSLEQDLSLDLLPRGMDSFKWASSTEKQPS 418

RESULT 11
US-10-408-765A-2211
; Sequence 2211, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 66088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2211
; LENGTH: 394
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-2211

```

Query Match 74.8%; Score 1538; DB 16; Length 394;
Best Local Similarity 93.3%; Pred. No. 7.2e-146;
Matches 294; Conservative 7; Mismatches 14; Indels 0; Gaps 0;

Qy 1 MWLPRVSTAVTALLAQTLLFLVSRPGSPSPAGGEARVHVLVLSWRSGSFGQLF 60
Db 1 MWLPRVSTAVTALLAQTLLFLVSRPGSPSPAGGEARVHVLVLSWRSGSFGQLF 60

Qy 61 NOHPDVFYLMPEPAWHVMTTLLSQSAAATLHMAVRDLVRSVFLCDMDVFDAYLPWRRLNLDL 120
Db 61 NOHPDVFYLMPEPAWHVMTTLLSQSAAATLHMAVRDLVRSVFLCDMDVFDAYLPWRRLNLDL 120

Qy 121 FQAVSRALCSPACSPACSAFFRGAISSSEAVCKPLCARQSFLLAREACRSYSHVVLKEVRFFN 180
Db 121 FQAVSRALCSPACSPACSAFFRGAISSKQDVCKTLCAQFPFLAREACRSYSHVVLKEVRFFN 180

Qy 181 LQVLYPLSDPALNLRVHVLVRDPAVLRSGEOTAKALARDNGIVLGTNGTWVEADPGLR 240
Db 181 LQVLYPLSDPALNLRVHVLVRDPAVLRSGEAGPTLARDNGIVLGTNGTWVEADPGLR 240

Qy 241 VREVCSHVRIABEATLKPPPLRGYRLVRFDLAREPLAEIRALYAFTGLSLTPGLE 300
Db 241 LIREVCSHVRIABEATLKPPPLRGYRLVRFDLAREPLAEIRALYAFTGLSLTPGLE 300

Qy 301 AWIHNITHGSGPGAR 315
Db 301 AWIHNITHGSGASQ 315

RESULT 12
US-09-816-825-2
; Sequence 2, Application US/09816825
; Patent No. US20010051370A1
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CON
; CURRENT APPLICATION NUMBER: US/09/816,825
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 09/045,284
; PRIOR FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-816-825-2

Query Match 49.6%; Score 1019; DB 9; Length 386;
Best Local Similarity 54.5%; Pred. No. 1.4e-93;
Matches 216; Conservative 45; Mismatches 107; Indels 28; Gaps 7;

Qy 14 LLLAQTLLFLVSRP-----GPSSPAGGEARVHVLVLSWRSGSFGQL 59
Db 2 LLPKMKLLFLVSQMAILALFFHMYSHNTSSLSMKQAQPERMHVVLVLSWRSGSFGQL 61

Qy 60 FNQHPDVFYLMPEPAWHVMTTLLSQSAAATLHMAVRDLVRSVFLCDMDVFDAYL-PWRRNLS 118
Db 62 FGQHPDVFYLMPEPAWHVMTFKQSTAWMLHMAVRDLIRAVFLCDMSVFDAYMEGPRQS 121

Qy 119 DLFOAVSRALCSPACSAFFRGAISSSEAVCKPLCARQSFLLAREACRSYSHVVLKEVRFF 178
Db 122 SLFQWNSRALCSPACDIIPQDEIIIPRAHCRLLCSQPPFVVEKACRSYSHVVLKEVRFF 181

Qy 179 FNLQVLYPLSDPALNLRVHVLVRDPAVLRSGEOTAKALARDNGIVLGTNGTWV-EADP 237
Db 182 FNLQSLYPLKDPSELNHLVHVLVRDPAVFRSRRTKGDLMIDSRIVMGQHEOKLKKEQD 241

Qy 238 GLRVVREVCCHVRIABEATLK-PPLFGRGRYRLVRFDLAREPLAEIRALYAFTGLSLT 296
Db 242 PYVMQVICSQLEIYK--TIQSLPKALQERYLLVRYEDLARAPVPAQTSRMFYEVGLLE 299

Qy 297 POLEAWHNITHGSGPGARREAFKTSRNALNVSOAWRHALPFAKIRRVQELCAGALQLL 356
Db 300 PHLQTVHNIITRGKMG--DHAFHTNARDALNVSOAWRWSLPYEKVSRLQKACGDAMNLL 357

RESULT 14
US-10-841-707-6
; Sequence 6, Application US/10841707
; Publication No. US20040202649A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Minoru

Db 242 PYVMQVICSQLEIYK--TIQSLPKALQERYLLVRYEDLARAPVPAQTSRMFYEVGLLE 299

Qy 297 POLEAWHNITHGSGPGARREAFKTSRNALNVSOAWRHALPFAKIRRVQELCAGALQLL 356
Db 300 PHLQTVHNIITRGKMG--DHAFHTNARDALNVSOAWRWSLPYEKVSRLQKACGDAMNLL 357

Qy 357 GYRPVYSEDEQRNLALDLVLPRLNGFTWASSTASH 392
Db 358 GYRHVRSEQRNLALDL-----STWTVPEQIH 386

RESULT 13
US-10-007-262-1
; Sequence 1, Application US/10007262
; Publication No. US20020164748A1
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Tangemann, Kirsten
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CIP
; CURRENT APPLICATION NUMBER: US/10/007,262
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/190,911
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-12
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 386
; TYPE: PRT
; ORGANISM: H. sapiens
US-10-007-262-1

Query Match 49.6%; Score 1019; DB 13; Length 386;
Best Local Similarity 54.5%; Pred. No. 1.4e-93;
Matches 216; Conservative 45; Mismatches 107; Indels 28; Gaps 7;

Qy 14 LLLAQTLLFLVSRP-----GPSSPAGGEARVHVLVLSWRSGSFGQL 59
Db 2 LLPKMKLLFLVSQMAILALFFHMYSHNTSSLSMKQAQPERMHVVLVLSWRSGSFGQL 61

Qy 60 FNQHPDVFYLMPEPAWHVMTTLLSQSAAATLHMAVRDLVRSVFLCDMDVFDAYL-PWRRNLS 118
Db 62 FGQHPDVFYLMPEPAWHVMTFKQSTAWMLHMAVRDLIRAVFLCDMSVFDAYMEGPRQS 121

Qy 119 DLFOAVSRALCSPACSAFFRGAISSSEAVCKPLCARQSFLLAREACRSYSHVVLKEVRFF 178
Db 122 SLFQWNSRALCSPACDIIPQDEIIIPRAHCRLLCSQPPFVVEKACRSYSHVVLKEVRFF 181

Qy 179 FNLQVLYPLSDPALNLRVHVLVRDPAVLRSGEOTAKALARDNGIVLGTNGTWV-EADP 237
Db 182 FNLQSLYPLKDPSELNHLVHVLVRDPAVFRSRRTKGDLMIDSRIVMGQHEOKLKKEQD 241

Qy 238 GLRVVREVCCHVRIABEATLK-PPLFGRGRYRLVRFDLAREPLAEIRALYAFTGLSLT 296
Db 242 PYVMQVICSQLEIYK--TIQSLPKALQERYLLVRYEDLARAPVPAQTSRMFYEVGLLE 299

Qy 297 POLEAWHNITHGSGPGARREAFKTSRNALNVSOAWRHALPFAKIRRVQELCAGALQLL 356
Db 300 PHLQTVHNIITRGKMG--DHAFHTNARDALNVSOAWRWSLPYEKVSRLQKACGDAMNLL 357

Qy 357 GYRPVYSEDEQRNLALDLVLPRLNGFTWASSTASH 392
Db 358 GYRHVRSEQRNLALDL-----STWTVPEQIH 386

RESULT 14
US-10-841-707-6
; Sequence 6, Application US/10841707
; Publication No. US20040202649A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Minoru

```
; APPLICANT: Yeh, Jiunn-Chern
; APPLICANT: Hiraoka, Nobuyoshi
; TITLE OF INVENTION: Identification of the Meca-79 Antigen
; TITLE OF INVENTION: and Related Methods of Treating L-Selectin-Mediated
; TITLE OF INVENTION: Conditions
; FILE REFERENCE: P-LJ 4149
; CURRENT APPLICATION NUMBER: US/10/841,707
; CURRENT FILING DATE: 2004-05-06
; PRIOR APPLICATION NUMBER: US/09/569,320A
; PRIOR FILING DATE: 2000-05-11
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-841-707-6

Query Match      49.5%; Score 1018; DB 16; Length 380;
Best Local Similarity 55.0%; Pred. No. 1.7e-93;
Matches 214; Conservative 44; Mismatches 103; Indels 28; Gaps 7;

Qy 21 LLLFLVSRP-----GPPSPAGGEARVHVLLVSSWSSGSSFGVQLFNOHPDV 66
Db 3 LLLFLVSRP-----GPPSPAGGEARVHVLLVSSWSSGSSFGVQLFNOHPDV 62
Qy 67 FYLMEPAHWVMTTSLQSSAATLHMVAVRDLVRSVFLCDMDVFDAYL-PWRRNLSDLFOWAV 125
Db 63 FYLMEPAHWVMTTSLQSSAATLHMVAVRDLVRSVFLCDMDVFDAYL-PWRRNLSDLFOWAV 122
Qy 126 SRALCSPACSAFPRGAISSBAVCKPLCARQSFTLAREACRSYSHVVLKVRFPFNLOVLY 185
Db 123 SRALCSPACSAFPRGAISSBAVCKPLCARQSFTLAREACRSYSHVVLKVRFPFNLOVLY 182
Qy 186 PLLSDPALNLRIVLVRDPRVAVRSRQTAALARDNGIVLGTNGTWV-EADPGLRVURE 244
Db 183 PLLSDPALNLRIVLVRDPRVAVRSRQTAALARDNGIVLGTNGTWV-EADPGLRVURE 242
Qy 245 VCRSHVRIAEAAATLKP-PPLRGRVRLVRFEDLAREPLAEIRALYAFGLSLTPOLEAWI 303
Db 243 ICQSQLEIYK--TIQSLPKALQERYLLVRYEDLARAPVAQTSRMVFEVGLFLEPHLQTWV 300
Qy 304 HNITHGSPGARREAFKTSRRNALNVSQAWRHALPFPAKIRRVQELCAGALQLLGYRPPVYS 363
Db 301 HNITHGSPGARREAFKTSRRNALNVSQAWRHALPFPAKIRRVQELCAGALQLLGYRPPVYS 358
Qy 364 EDEORNALDLVLRGLNGFTWASSTASH 392
Db 359 EDEORNALDLVLRGLNGFTWASSTASH 380

RESULT 15
US-10-427-631-11
; Sequence 11, Application US/10427631
; Publication No. US20030175923A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION; TANG, Y. Tom;
; APPLICANT: CORLEY, Neil C.; GUEGLER, Karl J.;
; APPLICANT: BAUGHN, Mariah R.; LAL, Preeti G.;
; APPLICANT: YUE, Henry; HILLMAN, Jennifer L.;
; APPLICANT: AZIMZAI, Yalda
; TITLE OF INVENTION: HUMAN TRANSFERASE PROTEINS
; FILE REFERENCE: PF-0592-1 DIV
; CURRENT APPLICATION NUMBER: US/10/427,631
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 09/786,240
; PRIOR FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: PCT/US99/20989
; PRIOR FILING DATE: 1999-09-09
; PRIOR APPLICATION NUMBER: US 60/172,220
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: US 60/155,248
; PRIOR FILING DATE: 1998-11-04
```

```
; PRIOR APPLICATION NUMBER: US 60/133,642
; PRIOR FILING DATE: 1999-05-11
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PERL Program
; SEQ ID NO 11
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030175923A1 2617407CD1
US-10-427-631-11

Query Match      47.7%; Score 981; DB 14; Length 386;
Best Local Similarity 54.2%; Pred. No. 9.4e-90;
Matches 215; Conservative 42; Mismatches 110; Indels 30; Gaps 9;

Qy 14 LLLAQTFLLFLVSRP-----GPPSPAGGEARVHVLLVSSWSSGSSFGVQL 59
Db 2 LLLPKMKLLLLFLVSRP-----GPPSPAGGEARVHVLLVSSWSSGSSFGVQL 61
Qy 60 FNOHPDVFYLMPEPAHWVMTTSLQSSAATLHMVAVRDLVRSVFLCDMDVFDAYL-PWRRNLS 118
Db 62 FNOHPDVFYLMPEPAHWVMTTSLQSSAATLHMVAVRDLVRSVFLCDMDVFDAYL-PWRRNLS 121
Qy 119 DLFQWAVSRALCSPACSAFPRGAISSBAVCKPLCARQSFTLAREACRSYSHVVLKVR 177
Db 122 DLFQWAVSRALCSPACSAFPRGAISSBAVCKPLCARQSFTLAREACRSYSHVVLKVR 180
Qy 178 FPNLOVLYPLSDPALNLRIVLVRDPRVAVRSRQTAALARDNGIVLGTNGTWV-EAD 236
Db 181 FPNLOVLYPLSDPALNLRIVLVRDPRVAVRSRQTAALARDNGIVLGTNGTWV-EAD 240
Qy 237 PGLRVVREVCRSYSHVRIAEAAATLKP-PPLRGRVRLVRFEDLAREPLAEIRALYAFGLSL 295
Db 241 PGLRVVREVCRSYSHVRIAEAAATLKP-PPLRGRVRLVRFEDLAREPLAEIRALYAFGLSL 298
Qy 296 TPQLEAWIHNIHITHGSPGARREAFKTSRRNALNVSQAWRHALPFPAKIRRVQELCAGALQL 355
Db 299 TPQLEAWIHNIHITHGSPGARREAFKTSRRNALNVSQAWRHALPFPAKIRRVQELCAGALQL 356
Qy 356 LGYRPPVYSEDEORNALDLVLRGLNGFTWASSTASH 392
Db 357 LGYRPPVYSEDEORNALDLVLRGLNGFTWASSTASH 386

Search completed: June 23, 2005, 09:27:57
Job time : 51.935 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2005, 08:33:38 ; Search time 8.95931 Seconds
(without alignments)
4242.027 Million cell updates/sec

Title: US-10-697-828-13
Perfect score: 2056
Sequence: 1 MMLPRVSSSTAVTALLLAQTF.....LPRGLNGFTWASSTASHPRN 395

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_79:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----------|---------------------|
| 1 | 656.5 | 31.9 | 484 | 2 JE0261 | N-acetylglucosamin |
| 2 | 618.5 | 30.1 | 484 | 2 JC7350 | N-acetylglucosamin |
| 3 | 615 | 29.9 | 486 | 2 JC7351 | N-acetylglucosamin |
| 4 | 532.5 | 25.9 | 458 | 2 A57337 | chondroitin 6-sulf |
| 5 | 128 | 6.2 | 388 | 2 G70729 | hypothetical prote |
| 6 | 118 | 5.7 | 307 | 2 E95934 | probable enzyme, C |
| 7 | 100.5 | 4.9 | 523 | 2 B83629 | probable ATP-bindi |
| 8 | 98 | 4.8 | 559 | 2 T03412 | malate synthase (E |
| 9 | 93 | 4.5 | 359 | 2 T16350 | hypothetical prote |
| 10 | 91.5 | 4.5 | 425 | 2 E83023 | 3-deoxy-D-manno-oc |
| 11 | 91.5 | 4.5 | 577 | 2 B75585 | probable long-chain |
| 12 | 91.5 | 4.5 | 617 | 2 F82744 | ferrous iron trans |
| 13 | 91.5 | 4.5 | 820 | 2 H86246 | hypothetical prote |
| 14 | 91 | 4.4 | 380 | 2 A75328 | phoH-related prote |
| 15 | 90.5 | 4.4 | 349 | 2 E75611 | glucosamine-fructo |
| 16 | 90.5 | 4.4 | 710 | 1 A46273 | mitochondrial inte |
| 17 | 90 | 4.4 | 931 | 2 AF3276 | ATP-dependent clp |
| 18 | 89.5 | 4.4 | 221 | 2 T50665 | hypothetical prote |
| 19 | 89.5 | 4.4 | 668 | 2 T01685 | crp1 protein - mai |
| 20 | 89.5 | 4.4 | 1489 | 2 S73015 | polyketide synthas |
| 21 | 89.5 | 4.4 | 3643 | 2 T36410 | probable polyketid |
| 22 | 89 | 4.3 | 571 | 2 AG3174 | acetolactate synth |
| 23 | 88.5 | 4.3 | 262 | 2 G95327 | Trm3 transposase f |
| 24 | 88.5 | 4.3 | 400 | 2 S24759 | probable transposa |
| 25 | 88.5 | 4.3 | 400 | 2 A42727 | probable transposa |
| 26 | 88.5 | 4.3 | 400 | 2 C95306 | Trm3 transposase [|
| 27 | 88.5 | 4.3 | 400 | 2 E95287 | Trm3 transposase [|
| 28 | 88.5 | 4.3 | 400 | 2 H95291 | Trm3 transposase [|
| 29 | 88.5 | 4.3 | 400 | 2 P95354 | Trm3 transposase [|

| | | | | | |
|----|------|-----|-----|----------|--------------------|
| 30 | 88.5 | 4.3 | 809 | 2 C71279 | probable ATP-depen |
| 31 | 88 | 4.3 | 370 | 2 A49360 | pyruvate dehydroge |
| 32 | 88 | 4.3 | 390 | 2 G82844 | cysteine synthase |
| 33 | 87.5 | 4.3 | 260 | 2 B95965 | probable two-compo |
| 34 | 87.5 | 4.3 | 467 | 2 E95981 | conserved hypotet |
| 35 | 87.5 | 4.3 | 547 | 1 S51475 | cytochrome P450 Cy |
| 36 | 87 | 4.2 | 327 | 2 T50744 | spheroidene monoox |
| 37 | 87 | 4.2 | 389 | 1 DEGP4 | pyruvate dehydroge |
| 38 | 87 | 4.2 | 532 | 2 S33758 | flavin-containing |
| 39 | 86.5 | 4.2 | 570 | 2 G95326 | probable acetolact |
| 40 | 86.5 | 4.2 | 856 | 2 F87316 | conserved hypotet |
| 41 | 86 | 4.2 | 528 | 2 T51119 | phytoene desaturas |
| 42 | 86 | 4.2 | 652 | 2 T45499 | probable DNA recom |
| 43 | 86 | 4.2 | 787 | 2 AB3207 | conserved hypotet |
| 44 | 85 | 4.1 | 293 | 2 C71267 | probable carboxyle |
| 45 | 85 | 4.1 | 327 | 2 S49619 | crtA protein - Rho |

ALIGNMENTS

RESULT 1
JE0261
N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) - human
C;Species: Homo sapiens (man)
C;Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 01-Mar-2002
C;Accession: JE0261
R;Uchimura, K.; Muramatsu, H.; Kaname, T.; Ogawa, H.; Yamakawa, T.; Fan, Q.W.; Mitsuoka, T.
J. Biochem. 124, 670-678, 1998
A;Title: Human N-Acetylglucosamine-6-O-sulfotransferase involved in the biosynthesis of f
A;Reference number: JE0261; MUID:98391845; PMID:9722682
A;Accession: JE0261
A;Molecule type: mRNA
A;Residues: 1-484 <UCH>
A;Cross-references: DDBJ:AB014679
C;Comment: This protein catalyzes the transfer of sulfer of sulfate from 3'-phosphoadenos
C;Superfamily: chondroitin 6-sulfotransferase
C;Keywords: sulfotransferase

| | | | | | |
|-----------------------|-----|--|--------------------|-------|-------------------------------------|
| Query Match | | 31.9% | Score 656.5; | DB 2; | Length 484; |
| Best Local Similarity | | 39.0% | Pred. No. 5.9e-50; | | |
| Matches | | 146; | Conservative | 69; | Mismatches 124; Indels 35; Gaps 10; |
| Qy | 33 | SPAG-GEARVHVLVLSWSRSGSVGQLEFNQHDVFLMEPAWHVWTTLSQGSAAHLMA | 91 | | |
| Db | 110 | APEGVGDKRHWYVFTTWRSQSGSFFGELFNQNPVEFLYEPVHWOKLYPGDAVSLQGA | 169 | | |
| Qy | 92 | VRDLVRSVFLCDMDVFDAYLP---WRNLSDL--FQWAVSRALCSPACSAFPRGAIS-- | 144 | | |
| Db | 170 | ARMILSALYRCDLSVLQYSPAGSGGNLTTLGFGAATNKVVCSSPLCPAYRKEVVGLV | 229 | | |
| Qy | 145 | SEAVCKPLCARQFTLAREACRSYHVLVKEVFFNQLVLYPLSLDPAALNLRIVHLVRDP | 204 | | |
| Db | 230 | DDRVCCK-CPQRLARFEBCRKRYTLVIGKRVFDDVAVLAPLLRDPALDPLKVIHLVRDP | 288 | | |
| Qy | 205 | RAVLSRBEQTAKALARDN-----GIVLGT--NGTWVAD-PGLRV | 241 | | |
| Db | 289 | RAVASSRIRSRHGLIRSLQVRSRDPRAHRMPFLEAAGHKJGAKGEGVGGPDYHALGA | 348 | | |
| Qy | 242 | VREVCVSHVRIABAATLKPPFLRGVRLRVRFEDLAREPLAEIRALYAFATGLSLTPOLEA | 301 | | |
| Db | 349 | MEVICHSMATLQTA-LQPPDWLQGHVLRVEDLVCEPVKTLRRVYDFVGLLVSPMEQ | 407 | | |
| Qy | 302 | WIHNITHGSGPGARRAFAKTTSSRNALNVSOAQRHALPFAKIRRVQELCAGALQLLGRVP | 361 | | |
| Db | 408 | FALNMTSGSGSSSK--PFVVSARNATQAANAWRTALTFOQIKQVEEFCYOPMAVLGVERV | 465 | | |
| Qy | 362 | YSEDEQRNALDLV | 375 | | |
| Db | 466 | NSPEEVKDLKSTLL | 479 | | |

RESULT 2
 JC7350
 N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) [validated] - mouse
 C;Species: Mus musculus (house mouse)
 C;Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 09-Jul-2004
 C;Accession: JC7350
 R;Uchimura, K.; Fasakhany, F.; Kadomatsu, K.; Matsukawa, T.; Yamakawa, N.;
 Biochem. Biophys. Res. Commun. 274, 291-296, 2000
 A;Title: Diversity of N-acetylglucosamine-6-O-sulfotransferases: Molecular cloning of a
 A;Reference number: JC7350; MUID:20374462; PMID:10913333
 A;Accession: JC7350
 A;Molecule type: mRNA
 A;Residues: 1-484 <UCH>
 A;Cross-references: UNIPROT:Q99NB0; UNIPROT:Q9EP78; DDBJ:AB040710
 C;Comment: This enzyme, having a type II transmembrane topology and N-linked glycosylation
 sulfate linkage formation. This enzyme is involved in synthesis of L-selectin ligand in
 C;Keywords: glycolysis; sulfate transport; sulfotransferase; transmembrane protein

Query Match 30.1%; Score 618.5; DB 2; Length 484;
 Best Local Similarity 38.0%; Pred. No. 1.3e-46;
 Matches 148; Conservative 56; Mismatches 130; Indels 55; Gaps 11;

QY 31 PSSP-----ACGEA----RVHVLVSSWRSGSFVGLFNOHPDVFLMEPAWHVWTTLSQ 82
 DB 82 PGSPGNLSAAGEAVTQEKQHYHATWRTGSSFLGELFNOHPDVFLYEPMHLWQALY 141

QY 83 GSAATLHMAVRDLVRSVFLCDMDVFDAY-----LPMRRLSD--LFQWAVSRALCS 131
 DB 142 GDAESLQALRDLMLRSFLRCDFSVLRLYAQDPGGERAPDSANLTTAMLFRWRTNKVICS 201

QY 132 PPACSAFPR-----GAISSEAVCKPFCARQSFLLAREACRSYSHVVLKEVRFNLOVLYP 186
 DB 202 PPLCPAAPARADVGLVEDKA-CESTCPPVSLRALEAECKRPVWVKDVLRLDLGLVLP 260

QY 187 LLSDPALNLRIVHLVRDPRAVLRSEQTAKALARDNGIVLGTN-----GTWVEA 235
 DB 261 LLRDPGLNLKVVQLFRDPRAVHNSRLKSRQLRESIQVLRQGRDHFHRLVLAHGVD 320

QY 236 DPG-----LRVREVCRSYHVRVIAEATLKPPFLGRYLRVRFEDLARE 279
 DB 321 RPPGQARALPSAPRADFFLTSALEVICAEWLRL-LFTRGAPAWLRRYLRVLYEDLVWQ 379

QY 280 PLAEIRALYFTGLSLTPQLAEWHNTHGSGPGARREAFKTSRRNALNVSOAWRHLP 339
 DB 380 PQQLRRLRFLSGRLAALDAFALNMTGAAYGADR-PFHLSARDAREAVHWRERLSQ 438

QY 340 AKIRRVQELCAGALQLLGYRVPVISEDEQR 368
 DB 439 EQVRQVETACAPAMRLLAY-PRSGDERDR 466

RESULT 3
 JC7351
 N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) [validated] - human
 C;Species: Homo sapiens (man)
 C;Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 09-Jul-2004
 C;Accession: JC7351
 R;Uchimura, K.; Fasakhany, F.; Kadomatsu, K.; Matsukawa, T.; Yamakawa, N.;
 Biochem. Biophys. Res. Commun. 274, 291-296, 2000
 A;Title: Diversity of N-acetylglucosamine-6-O-sulfotransferases: Molecular cloning of a
 A;Reference number: JC7350; MUID:20374462; PMID:10913333
 A;Accession: JC7351
 A;Molecule type: mRNA
 A;Residues: 1-486 <UCH>
 A;Cross-references: UNIPROT:Q75667; DDBJ:AB040711
 C;Comment: This enzyme, having a type II transmembrane topology and N-linked glycosylation
 sulfate linkage formation. This enzyme is involved in synthesis of L-selectin ligand in
 C;Keywords: glycolysis; sulfate transport; sulfotransferase; transmembrane protein

Query Match 29.9%; Score 615; DB 2; Length 486;
 Best Local Similarity 38.0%; Pred. No. 2.7e-46;
 Matches 144; Conservative 53; Mismatches 128; Indels 54; Gaps 8;

QY 29 RPPSPAGGEA----RVHVLVSSWRSGSFVGLFNOHPDVFLMEPAWHVWTTLSQGS 84
 DB 86 PSNLSGAVGEAVSRKQHYHATWRTGSSFLGELFNOHPDVFLYEPMHLWQALYPGD 145

QY 85 AATLHMAVRDLVRSVFLCDMDVFDAY-----LPMRRLSD--LFQWAVSRALCSPP 133
 DB 146 AESLQALRDLMLRSFLRCDFSVLRLYAQDPGPAARADPTANLTTAALFRWRTNKVICSPP 205

QY 134 ACSAPPRG-----AISSEAVCKPFCARQSFLLAREACRSYSHVVLKEVRFNLOVLYPLS 189
 DB 206 LCPGAPRARAEVGLVEDTACERSCPPVAIRALEAECKRPVWVKDVLRLDLGLVLP 265

QY 190 DPALNLRIVHLVRDPRAVLRSEQTAKALARDNGIVLGTN-----GTWVEADPG 238
 DB 266 DPGLNLKVVQLFRDPRAVHNSRLKSRQLRESIQVLRQGRDHFHRLVLAHGVARPG 325

QY 239 -----LRVREVCRSYHVRVIAEATLKPPFLGRYLRVRFEDLARE 279
 DB 326 QOSRALPAAPRADFFLTGALVICAEWLRLDLFARGA-----PAWLRRYLRVLYEDLV 381

QY 280 PLAEIRALYFTGLSLTPQLAEWHNTHGSGPGARREAFKTSRRNALNVSOAWRHLP 339
 DB 382 PRAQLRRLRFLSGRLAALDAFALNMTGAAYGADR-PFHLSARDAREAVHWRERLSR 440

QY 340 AKIRRVQELCAGALQLLGY 358
 DB 441 EQVRQVETACAPAMRLLAY 459

RESULT 4
 AS7397
 Chondroitin 6-sulfotransferase (EC 2.8.2.17) precursor - chicken
 C;Species: Gallus gallus (chicken)
 C;Date: 23-Feb-1996 #sequence_revision 23-Feb-1996 #text_change 09-Jul-2004
 C;Accession: AS7397
 J;Fukuta, M.; Uchimura, K.; Nakashima, K.; Kato, M.; Kimata, K.; Shinomura, T.; Habuchi,
 J. Biol. Chem. 270, 18575-18580, 1995
 A;Title: Molecular cloning and expression of chick chondrocyte chondroitin 6-sulfotransferase
 A;Reference number: AS7397; MUID:95355490; PMID:7629189
 A;Accession: AS7397
 A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 1-458 <FUK>
 A;Cross-references: UNIPROT:Q92179; GB:D49915; NID:g971262; PIDN:BAA08655.1; PID:g971263
 C;Superfamily: chondroitin 6-sulfotransferase
 C;Keywords: sulfotransferase

Query Match 25.9%; Score 532.5; DB 2; Length 458;
 Best Local Similarity 34.6%; Pred. No. 4.9e-39;
 Matches 134; Conservative 63; Mismatches 157; Indels 33; Gaps 10;

QY 5 RVSSTAVTALLAQTFLLFLVSRPGSPAGGEARVHVLVSSWRSGSFVGLFNOHP 64
 DB 80 RQLDTFAPSQLATRLRNVTQLAGELGIAAP--EPRRHVLLMATRTGSSFVGEFNOQG 136

QY 65 DVFLYMEPAWHV--WTTLSQCSAATLHMAV--RDLVRSVFLCDMDVFDAYL---PWRNL 117
 DB 137 NIFLYFEPLWHIERVTVPFGANAVGALVYRDVLQQLLCDLYLEFSISPAPPEHLT 196

QY 118 SDLFQWAVSRALCSPPACSAFPRGAISSEAVCKPFCARQSFLLAREACRSYSHVVLKEVR 177
 DB 197 AALFRGSSHSICEPEVCTPSLKKVFEKYCKNRCGPNLITLAAEACRRKQHMALKTVR 256

QY 178 FPNLQVLYPLSDPALNLRIVHLVRDPRAVLRSEQTAKALARDNGIVLGTNGTW----V 233
 DB 257 IRQLEFLQPLAEDPRDLRIIQLVRDPRAVLSRVMAFS-----GKYESNKKWAA 306

QY 234 EADPGLRV-----VREVCRSYHVRVIAEATLKPPFLGRYLRVRFEDLAREPLAEIRALY 288
 DB 307 EGEPALQEDVORLGNCS-IRLSAELGLRQPRWLGRVYMLRYEDVARAPLRAKALEMY 365

QY 289 APTGLSLTPQLAEWHNTHGSGPGARREAFKTSRRNALNVSOAWRHLPFAKIRRVQEL 348

chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.M.; Marques, M.V.; Martins, E.Y.;
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
., P.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.;
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasaki,
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira,
M.; Tsuchiko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z

A;Reference number: A59328
A;Contents: annotation
A;Genetics:
C:Gene: XP0933
C:Superfamily: ferrous iron transport protein B; translation elongation factor Tu homolog

Query Match 4.5%; Score 91.5; DB 2; Length 617;
Best Local Similarity 25.7%; Pred. No. 6.2;
Matches 47; Conservative 24; Mismatches 73; Indels 39; Gaps 7;

QY 11 VTALLAQFLLLPLVSRPSPSPAGGEARVHVLVLSWRSRSSVGVQLFNQHPDVFYLM 70
DB 456 VGGIILALITLLWFLAAPPSPGVTGPAIY-----SPAGRICHVWAGFPAPL 504
QY 71 EPAHWYMTTILSQSAAATLHMAVDLVRSVFLC----DMDFDAYLPWRNLSDLFO--WA 124
DB 505 GFSQWQICIALIPGLAAR-EVAVSSLATVYALSASNEEMAV-----RALSPLIQDGWS 555
QY 125 VSRALCSPPACSAAPPRAIGAISSAEVCKPLCARQSTFLAREACRSYSHVULKEVFFNLQVL 184
DB 556 LATAL-----SLLVWVIYAPMCISTLATIKRET-NSWKHMAMTAGYLFALAYL 602
QY 185 YPL 187
DB 603 AAL 605

RESULT 13
H86246
hypothetical protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: H86246
R;Theologits, A.; Ecker, J.R.; Palm, C.J.; Federpiet, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Huizar, L.
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.;
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luroe, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I.
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A;Accession: H86246
A;Status: Preliminary
A;Molecule type: DNA
A;Residues: 1-820 <STO>
A/Cross-references: UNIPROT: Q9SXB4; GB: AE005172; NID: g5734728; PIDN: AAD49893.1.; GSPDB: GN000141.
C;Genetics:
 A;Map position: 1
C;Superfamily: S-receptor kinase; protein kinase homology; S-locus-specific glycoprotein

| | | | | |
|---------|---|-------------------|----------------|--------------------------|
| | Query Match | 4.5%; Score 91.5; | DB 2; | Length 820; |
| | Best Local Similarity | 22.8%; | Pred. No. 8.9; | |
| Matches | 78; Conservative | | Matches | 91; Indels 135; Gaps 21; |
| QY | 2 WLP-----RVSSTAATLLLAQTFLLLFVLRPSSPAGGARHVHLVLSS---- | 48 | | |
| Db | -----ITSWKPSDPSPG-SYTAAALVLAAYPELF 203 | | | |
| QY | 49 -----WISGSFVGLENOHPDVF-----YLMEPAHWHTTLSQGSAATLHM 90 | | | |
| Dd | -----PWNGQMENGLPDVIACVFLRYFNDDTGSVTMYSANDSLTRY 262 | | | |
| QY | 91 AVRDLVRSFVLCDMDVDFAIPLW-----RRNLTS-----DLFOHVASRLCSPP---PA 134 | | | |

Db 263 FYMDYRGSVIRRD-----WSETRRNTWVGLOVPATECDNYRRCGEFATCNPRKNPL 313
QY 135 CSAF-----PRGAI-----SSEAVCK---PL-CARQSFTLAREACRSYSHVVLKEVRFFNL 161
Db 314 CSIRGFRPNLLEWNGNWSGGCTRRVPLOCRQNNNGS----- 353
QY 182 QVLYPLLSDPALNLRIVHL-----VRDPRAVLRSEQTAKALARDNGIVLGT---NG 230
Db 354 -----ADGFLRLRMKLPDFARRSEASEPEC-LRTCLQTCSCIAAAHGLGYGCHWNG 405
QY 231 TWVEADPGLRVREVCRS-----HVRIA--EAATLKPPPLRG 266
Db 406 SLVDS-----QELSGASGLDLYRLAHSEBIKTDRPILIG 440

RESULT 14
A75328
PhoH-related protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: A75328
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: A75328
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-380 <WHI>
A:Cross-references: UNIPROT:Q9RSY1; GB:AE002038; GB:AE00513; NID:G6459790; PIDN:AAF1154
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR1988
A:Map position: 1
C:Superfamily: phosphate starvation-induced protein

Query Match 4.4%; Score 91; DB 2; Length 380;
Best Local Similarity 26.5%; Pred. No. 3.7;
Matches 68; Conservative 25; Mismatches 104; Indels 60; Gaps 14;

QY 124 AVSRALCSPACSAFPGRGAISSSEAVCKPLCARQSFTLARE----ACRSYSHVVLKEVRFF 179
Db 2 ALSECLFSP-----FPPRLGED--CAPAGRTGDKLTDQTNIAPEADQTVTAITLNL 55
QY 180 NLQVLYPLLSDPALNLRIVHLRDPRAVLRSEBOT-AKALARDNGIVLGTNGTWE-ADP 237
Db 56 DQREAVALLGANDANLR-----RMRELTRAKLIARTGETVTITGDAADVEGAER 103
QY 238 GLRVREVCRSHVRIAEATLKPPPLRGVRLVRFEDLAREPLAEIRALYAFTGLSL-- 295
Db 104 MVRDALDVRS-----GGELTDFSLRS-----ARLSSEGRSLAAETQV-----NGLTLPR 149
QY 296 -----TPQLEAWI-----HNITHSGPGGARREAFKTSRRNALNVSOAWRHLPFAKIRRV 345
Db 150 GLRKPFTGQKLYLDLINESDITFGVPAG---TCKTYMAVAMAV-----QALAKAKVKRI 201
QY 346 --QELCAGALQLLGYRP 360
Db 202 ILTRPAVEAGEKGLGFLP 218

RESULT 15
E75611
glucosamine-fructose-6-phosphate aminotransferase-related protein - Deinococcus radiodur
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: E75611
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: E75611
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-349 <WHI>
A:Cross-references: UNIPROT:Q9RYZ9; GB:AE001862; GB:AE001825; NID:G6460468; PIDN:AAF1221
A:Experimental source: strain R1
C:Genetics:
A:Gene: DRA0154
A:Map position: 2

Query Match 4.4%; Score 90.5; DB 2; Length 349;
Best Local Similarity 24.1%; Pred. No. 3.6;
Matches 78; Conservative 34; Mismatches 133; Indels 79; Gaps 15;

QY 33 SPAGGEARVHVLVLSWSRSGSFVQGLFNOHPDVFYLM-----EPAM-----H 75
Db 2 TPDASAPAEPLMLREMRAPAVVRELLNENQDAVARLANAIRERQPAVAVTTIARGSSDH 61
QY 76 VMTTLSQSGAATLHMAVRDLVRSVFLCDMDVFDAYLPMWRNLSDLFQWAVSRALCSPPPAC 135
Db 62 ACTVLKYALETQSLPVASLGPSVH-----TLYGA-----RUDLAGALVIAVSQSGAGPDVV 113
QY 136 S-----APPRGAISSEAV---CKPLCARQSFTL-----AREACRSYSHVVLKEVRFFN 180
Db 114 ENVRMAREGALTVALVNVNVEDSFLAEAEFTLPLRCGPEKAVAATKSY----LASLCAF- 168
QY 181 LQVLYPLLSDPALNLRIVHLRDPRAVLRSEBOTAKALA-----RONGIVLGTNGTWEA 235
Db 169 LPVLAELTGDEALK-RALDALPEQLAHTLTLEDAAHAEHLAAGRYRFADNLLILARGHYGVA 227
QY 236 DPLRVVREVCRSHVRIAEATLKPPPLRGVRLV-----RPE 274
Db 228 QEAALKKETCGIHAEAYSAA-----EFSGPKRLLAEGLLPLGFSADAANWEATRQAYD 282
QY 275 DLAREPLAEIRALYAFTGLSL-TP 297
Db 283 DL-RAAGADLRLLGPVAGADLPTP 305

Search completed: June 23, 2005, 08:50:17
Job time : 10.9593 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2005, 08:33:13 ; Search time 43.3472 Seconds
(without alignments)
4666.302 Million cell updates/sec

Title: US-10-697-828-13

Perfect score: 2056

Sequence: 1 MWLPRVSSTAVTALLAQTF.....LPRGLNGFTWASSTASHPRN 395

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

UniProt_03:*

1: uniprot_sprot:*

2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|----------|---------------------|
| 1 | 2056 | 100.0 | 395 | 2 Q9GZX3 | Q9GZX3 homo sapien |
| 2 | 1729.5 | 84.1 | 390 | 2 Q7LCN3 | Q7LCN3 homo sapien |
| 3 | 1729.5 | 84.1 | 411 | 2 Q9GZS9 | Q9GZS9 homo sapien |
| 4 | 1708 | 83.1 | 395 | 2 Q9QUF4 | Q9QUF4 mus musculus |
| 5 | 1177 | 57.2 | 392 | 2 Q6GNJ9 | Q6GNJ9 xenopus lae |
| 6 | 1024 | 49.8 | 386 | 2 Q8NCQ5 | Q8NCQ5 homo sapien |
| 7 | 1019 | 49.6 | 386 | 2 Q9Y5R3 | Q9Y5R3 homo sapien |
| 8 | 1012 | 49.2 | 370 | 2 Q8IV46 | Q8IV46 homo sapien |
| 9 | 984 | 47.9 | 388 | 2 Q8RI11 | Q8RI11 mus musculus |
| 10 | 977 | 47.5 | 388 | 2 Q9W0B5 | Q9W0B5 mus musculus |
| 11 | 663 | 32.2 | 483 | 2 Q794G9 | Q794G9 mus musculus |
| 12 | 663 | 32.2 | 530 | 2 Q88276 | Q88276 mus musculus |
| 13 | 660 | 32.1 | 530 | 2 Q80WV3 | Q80WV3 mus musculus |
| 14 | 657.5 | 32.0 | 483 | 2 Q9UEB5 | Q9UEB5 homo sapien |
| 15 | 657.5 | 32.0 | 530 | 2 Q9Y4C5 | Q9Y4C5 homo sapien |
| 16 | 622.5 | 30.3 | 484 | 2 Q9EP78 | Q9EP78 m n-acetyl |
| 17 | 619.5 | 30.1 | 484 | 2 Q99NR0 | Q99NR0 mus musculus |
| 18 | 615 | 29.9 | 486 | 2 Q75667 | Q75667 homo sapien |
| 19 | 614 | 29.9 | 486 | 2 Q9NS84 | Q9NS84 homo sapien |
| 20 | 593 | 28.8 | 420 | 2 Q6DBY9 | Q6DBY9 brachydanio |
| 21 | 593 | 28.8 | 485 | 2 Q8XQ88 | Q8XQ88 rattus norv |
| 22 | 578.5 | 28.1 | 479 | 2 Q7LGC8 | Q7LGC8 homo sapien |
| 23 | 576.5 | 28.0 | 479 | 2 Q75099 | Q75099 homo sapien |
| 24 | 569 | 27.7 | 472 | 2 Q88199 | Q88199 mus musculus |
| 25 | 564.5 | 27.5 | 411 | 2 Q9EQC0 | Q9EQC0 mus musculus |
| 26 | 555.5 | 27.0 | 411 | 2 Q43916 | Q43916 homo sapien |
| 27 | 555 | 27.0 | 474 | 2 Q9QZ12 | Q9QZ12 rattus norv |
| 28 | 532.5 | 25.9 | 458 | 1 C6ST | Q92179 gallus gall |
| 29 | 463.5 | 22.5 | 441 | 2 Q93403 | Q93403 torpedo cal |
| 30 | 393.5 | 19.1 | 257 | 2 Q79415 | Q79415 mus musculus |
| 31 | 347.5 | 16.9 | 225 | 2 Q6RY62 | Q6RY62 cavia porce |

32 288.5 14.0 304 2 Q677T4 Q677T4 lymphocysti
33 270.5 13.2 486 2 Q9VMC3 Q9VMC3 drosophila
34 270 13.1 345 2 Q7Q1F9 Q7Q1F9 anopheles g
35 193 9.4 120 2 Q9DOK5 Q9DOK5 mus musculus
36 190 9.2 119 2 Q95JA8 Q95JA8 oryctolagus
37 169.5 8.2 363 2 Q9VMC4 Q9VMC4 drosophila
38 163.5 8.0 363 2 Q8MZD1 Q8MZD1 drosophila
39 133 6.5 303 2 Q93JE6 Q93JE6 streptomyce
40 128 6.2 388 1 Q967 MYCTU P64963 mycobacteri
41 128 6.2 388 1 YM90 MYCBO P64964 mycobacteri
42 123.5 6.0 1222 2 Q81ZU8 Q81ZU8 homo sapien
43 118 5.7 307 2 Q92VG4 Q92VG4 rhizobium m
44 114.5 5.6 274 2 Q8PRA0 Q8PRA0 xanthomonas
45 113.5 5.5 496 2 Q9EX08 Q9EX08 streptomyce

ALIGNMENTS

RESULT 1
Q9GZX3 PRELIMINARY; PRT; 395 AA.
ID Q9GZX3
AC Q9GZX3; 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 25-OCT-2004 (Tremblrel. 28, Last annotation update)
DE N-acetylglucosamine 6-O-sulfotransferase GST-4beta (Corneal N-
DE acetylglucosamine-6-O-sulfotransferase) (Carbohydrate N-
DE acetylglucosamine 6-O) sulfotransferase 6).
GN Name=GST4beta; Synonyms=CHST6;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21096027; PubMed=1181564; DOI=10.1093/glycob/11.1.75;
RA Hemmerich S., Lee J.-K., Bhakta S., Bistrup A., Ruddie N.R.,
RA Rosen S.D.;
RT "Chromosomal localization and genomic organization of the galactose/N-
RT acetylglucosamine/N-acetylglucosamine 6-O-sulfotransferase gene
RT family";
RL Glycobiology 11:75-87(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20472330; PubMed=11017086; DOI=10.1038/79987;
RA Akama T.O., Nishida K., Nakayama J., Watanabe H., Ozaki K.,
RA Nakamura T., Dota A., Kawasaki S., Inoue Y., Maeda N., Yamamoto S.,
RA Fujiwara T., Thonar E.J., Kinoshita S., Tanigami A.,
RA Fukuda M.N.;
RT "Macular corneal dystrophy type I and type II are caused by distinct
RT mutations in a new sulphotransferase gene";
RL Nat. Genet. 26:237-241(2000).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=22389257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,

```

RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RA Director MGC Project;
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF280086; AAC48244.1; -.
DR EMBL; AF219991; AAG26327.1; -.
DR EMBL; BC074883; AAH74883.1; -.
DR EMBL; BC074834; AAH74834.1; -.
DR EMBL; AF219990; AAC26325.1; -.
DR Genew; HGNC:6938; CHST6.
DR GO; GO:0005794; C:Golgi apparatus; TAS.
DR GO; GO:0001517; F:N-acetylglucosamine 6-O-sulfotransferase ac...; TAS.
DR GO; GO:0006044; P:N-acetylglucosamine metabolism; IC.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
KW Transferase.
SQ SEQUENCE 395 AA; 44098 MW; 433CA60248A48F67 CRC64;

Query Match 100.0%; Score 2056; DB 2; Length 395;
Best Local Similarity 100.0%; Pred. No. 1.6e-165;
Matches 395; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MWLPRVSSSTAVTALLAQTFLLFLVSRPGSPAGGEARVHVVLSSWRSGSSFGVQLF 60
DB 1 MWLPRVSSSTAVTALLAQTFLLFLVSRPGSPAGGEARVHVVLSSWRSGSSFGVQLF 60
QY 61 NQHPDVFYLMPEPAWHVWTTLSQSSAATLHMVRLVRSVFLCDMDVFDAYLPWRNLSL 120
DB 61 NQHPDVFYLMPEPAWHVWTTLSQSSAATLHMVRLVRSVFLCDMDVFDAYLPWRNLSL 120
QY 121 FQWAVSRALCSPACSAFPRGAISSBAVKPLCARQSFLLAREACRSYSHVVLKEVRFF 180
DB 121 FQWAVSRALCSPACSAFPRGAISSBAVKPLCARQSFLLAREACRSYSHVVLKEVRFF 180
QY 181 LQVLYPLSDPALNLRIVLVRDPAVLRSREQTAKALARDNGIVLGTNGTWVADPGLR 240
DB 181 LQVLYPLSDPALNLRIVLVRDPAVLRSREQTAKALARDNGIVLGTNGTWVADPGLR 240
QY 241 VREVCRSYHRIAFAEATLKPPPLRGYRLVRPDLAREPLAEIRALYFTGLSLTPQLE 300
DB 241 VREVCRSYHRIAFAEATLKPPPLRGYRLVRPDLAREPLAEIRALYFTGLSLTPQLE 300
QY 301 AWIHNITHGSGPGARREAFKTSRNALNVSQAWRHALPFAKIRRVQELCAGALQLIGYRP 360
DB 301 AWIHNITHGSGPGARREAFKTSRNALNVSQAWRHALPFAKIRRVQELCAGALQLIGYRP 360
QY 361 VYSEDEQRNALDLVLPRLNGFTWASSTASHPRN 395
DB 361 VYSEDEQRNALDLVLPRLNGFTWASSTASHPRN 395

RESULT 2
Q7LCN3 PRELIMINARY; PRT; 390 AA.
AC Q7LCN3;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE N-acetylglucosamine 6-O-sulfotransferase.
GN Name-I-GlcNAc-6-ST;
OS Homo sapiens (Human);
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Intestine;
RX MEDLINE=99423499; PubMed=10491328; DOI=10.1006/bbrc.1999.1324;

```

```

RA Lee J.K., Bhakta S., Rosen S.D., Hemmerich S.;
RT "Cloning and characterization of a mammalian N-acetylglucosamine-6-
RT sulfotransferase that is highly restricted to intestinal tissue.";
RL Biochem. Biophys. Res. Commun. 263:543-549(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Intestine;
RA Lee J.-K., Bhakta S., Rosen S.D., Hemmerich S.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF176839; AAD56001.1; -.
DR EMBL; AF176838; AAD56000.1; -.
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
KW Transferase.
SQ SEQUENCE 390 AA; 43832 MW; 132D1C3B6BE38179 CRC64;

Query Match 84.1%; Score 1729.5; DB 2; Length 390;
Best Local Similarity 85.8%; Pred. No. 7e-138;
Matches 333; Conservative 17; Mismatches 37; Indels 1; Gaps 1;

QY 1 MWLPRVSSSTAVTALLAQTFLLFLVSRPGSPAGGEARVHVVLSSWRSGSSFGVQL 59
DB 1 MWLPRFSSKTIVTLILAQTTCLLFIISRPGSPAGGEDRVHVVLSSWRSGSSFLGOL 60
QY 60 FNOHPDVFYLMPEPAWHVWTTLSQSSAATLHMVRLVRSVFLCDMDVFDAYLPWRNLSL 119
DB 61 FSOHPDVFYLMPEPAWHVWTTLSQSSAATLHMVRLVRSVFLCDMDVFDAYLPWRNLSA 120
QY 120 LFOWAVSRALCSPACSAFPRGAISSBAVKPLCARQSFLLAREACRSYSHVVLKEVRFF 179
DB 121 FENWATSALCSPACSAFPRGTISKQDVCKTLCTRQFSLAREACRSYSHVVLKEVRFF 180
QY 180 NLQVLYPLSDPALNLRIVLVRDPAVLRSREQTAKALARDNGIVLGTNGTWVADPGLR 239
DB 181 NLQVLYPLSDPALNLRIVLVRDPAVLRSREAGPIARDNGIVLGTNGKWEADPHL 240
QY 240 RVVRCRSYHRIAFAEATLKPPPLRGYRLVRPDLAREPLAEIRALYFTGLSLTPOL 299
DB 241 RLIREVCRSHVRIAFAEATLKPPPLRGYRLVRPDLAREPLAEIRALYFTGLTLTPOL 300
QY 300 EAWIHNITHGSGPGARREAFKTSRNALNVSQAWRHALPFAKIRRVQELCAGALQLIGYR 359
DB 301 EAWIHNITHGSGIGKPIEAFHTSSRNARNVSQAWRHALPFTKILRVQEVQECAGALQLIGYR 360
QY 360 VYSEDEQRNALDLVLPRLNGFTWAS 387
DB 361 VYVSADQQRDLTLDVLPRLGPDHFSWAS 388

RESULT 3
Q9GZS9 PRELIMINARY; PRT; 411 AA.
AC Q9GZS9; Q9UB3;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Intestinal GlcNAc-6-sulfotransferase (Intestinal N-acetylglucosamine-6-O-sulfotransferase).
GN Name=CHST5;
OS Homo sapiens (Human);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20472330; PubMed=11017086; DOI=10.1038/79987;
RA Akama T.O., Nishida K., Nakayama J., Watanabe H., Ozaki K.,
RA Nakamura T., Dota A., Kawasaki S., Inoue Y., Maeda N., Yamamoto S.,
RA Fujiwara T., Thonar E.J., Shimomura Y., Kinoshita S., Tanigami A.,
RA Fukuda M.N.;
RX "Macular corneal dystrophy type I and type II are caused by distinct

```

```

RT mutations in a new sulphotransferase gene.";
RL Nat. Genet. 26:237-241(2000).
DR EMBL; AF246718; AAC28023.1; -.
DR EMBL; AF219991; AAC26326.1; -.
DR Genew; HGNC:1973; CHST5.
DR GO; GO:0008146; F:sulphotransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000863; Sulphotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
KW Transferase.
SQ SEQUENCE 411 AA; 46160 MW; 97642D54BE926E06 CRC64;

Query Match      84.1%; Score 1729.5; DB 2; Length 411;
Best Local Similarity 85.8%; Pred. No. 7.4e-138;
Matches 333; Conservative 17; Mismatches 37; Indels 1; Gaps 1;

QY 1 MWLPVSVSTAVTALLAQ-TFLLLVSRGPPSPAGGEARVHVLVLSWRSGSFGQL 59
DB 22 MWLPVSVSTAVTALLAQ-TLLFIISRGPPSPAGGEDRVHVLVLSWRSGSFGQL 81
QY 60 FNOHPDVFLMEPAWHVWTTLSQGSAAATLHMVARDLVRSVFLCDMDVDFDAYLPMRRNLSL 119
DB 82 FSOHPDVFLMEPAWHVWTTLSQGSAAATLHMVARDLVRSVFLCDMDVDFDAYMPSRNLSA 141
QY 120 LFQWAVSRALCSPSPACSAFPRGAISSEAVCKPLCAROSFTLAREACRSYSHVVLKEVRFF 179
DB 142 FFWWATSRALCSPSPACSAFPRGTISKQDVCKTLCTROPFSLAREACRSYSHVVLKEVRFF 201
QY 180 NLOVLYPLSDPALNLRIVHLVRDPRAVLRSREQTAKALARDNGIVLGTNGTWVEADPGL 239
DB 202 NLOVLYPLSDPALNLRIVHLVRDPRAVLRSREQAAGPILARDNGIVLGTNGTWVEADPHL 261
QY 240 RVREVCRSVHRIAEEATLPPPLRGYRLVRFPEDLAREPLAEIRALYAFGTGLSLTPQL 299
DB 262 RLREVCRSVHRIAEEATLPPPLRGYRLVRFPEDLAREPLAEIRALYAFGTGLSLTPQL 321
QY 300 EAWIHNTTHSGGPGARREAFKTSRNALNVSQAWRHALPFAKIRRVQELCAGALQLGYR 359
DB 322 EAWIHNTTHSGGPGARREAFKTSRNALNVSQAWRHALPFAKIRRVQELCAGALQLGYR 381
QY 360 PVYSEDEQRNALDLVLPRLNGFTWAS 387
DB 382 PVYSAQQRDLTLVLPRLGPDHFSWAS 409

RESULT 4
Q9QUP4 ID Q9QUP4 PRELIMINARY; PRT; 395 AA.
AC Q9QUP4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 03-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE N-acetylglucosamine 6-O-sulphotransferase.
GN Name=Chst5; Synonyms=I-GlcNAC-6-ST;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=C57Bl/6; TISSUE=Intestine;
RP MEDLINE=99423499; PubMed=10491328; DOI=10.1006/bbrc.1999.1324;
RX Lee J.K., Bhakta S., Rosen S.D., Hemmerich S.;
RA "Cloning and characterization of a mammalian N-acetylglucosamine-6-
RT sulfotransferase that is highly restricted to intestinal tissue.";
RL Biochem. Biophys. Res. Commun. 263:543-549(1999).
RN [2]
RN SEQUENCE FROM N.A.
RP STRAIN=C57Bl/6; TISSUE=Intestine;
RC Lee J.-K., Bhakta S., Rosen S.D., Hemmerich S.;
RA Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF176841; AAD56003.1; -.
DR EMBL; AF176840; AAD56002.1; -.

```

```

DR MGD; MGI:1931825; Chst5.
DR GO; GO:0005615; C:extracellular space; TAS.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
KW Transferase.
SQ SEQUENCE 395 AA; 44537 MW; 3FDF71B43ED3B3BE CRC64;

Query Match      83.1%; Score 1708; DB 2; Length 395;
Best Local Similarity 83.0%; Pred. No. 4.7e-136;
Matches 328; Conservative 24; Mismatches 43; Indels 0; Gaps 0;

QY 1 MWLPVSVSTAVTALLAQ-TFLLLVSRGPPSPAGGEARVHVLVLSWRSGSFGQLF 60
DB 1 MRLPREFSSTVMLSLMLVQGITLVLSRQVPSPAGLGERVHVLVLSWRSGSFGQLF 60
QY 61 NOHPDVFLMEPAWHVWTTLSQGSAAATLHMVARDLVRSVFLCDMDVDFDAYLPMRRNLSL 120
DB 61 SQHPDVFLMEPAWHVWTTLSQGSAPALHMVARDLVRSVFLCDMDVDFDAYLPMRRNLSL 120
QY 121 FQWAVSRALCSPSPACSAFPRGAISSEAVCKPLCAROSFTLAREACRSYSHVVLKEVRFF 180
DB 121 FQWAVSRALCSPSPVCEAFARGNISSEVCKPLCATRPFGLAQEACRSYSHVVLKEVRFF 180
QY 181 LOVLYPLSDPALNLRIVHLVRDPRAVLRSREQTAKALARDNGIVLGTNGTWVEADPGLR 240
DB 181 LOVLYPLSDPALNLRIVHLVRDPRAVLRSREQTAKALARDNGIVLGTNGTWVEADPRLR 240
QY 241 VREVCRSVHRIAEEATLPPPLRGYRLVRFPEDLAREPLAEIRALYAFGTGLSLTPQL 300
DB 241 VNEVCRSVHRIAEEALHKKPPPLQDRYRLVRFPEDLAREPLTVIRELYAFGTGLTLTQLQ 300
QY 301 AWIHNTTHSGGPGARREAFKTSRNALNVSQAWRHALPFAKIRRVQELCAGALQLGYR 360
DB 301 TWIHNTTHSGGPGARREAFKTSRDALSVSQAWRHLPFAKIRRVQELCGALQLGYRS 360
QY 361 VYSEDEQRNALDLVLPRLNGFTWAS 395
DB 361 VHSELEQRDLTLVLPRLGMDSPKASSTKQPES 395

RESULT 5
Q6GN39 ID Q6GN39 PRELIMINARY; PRT; 392 AA.
AC Q6GN39;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE MGC83048 protein.
GN Name=MGC83048;
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OC NCBI_TaxID=8355;
RN [1]
RN SEQUENCE FROM N.A.
RP TISSUE=Embryo.
RC MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RX Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahney J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,

```


RX MEDLINE=21096027; PubMed=1181564; DOI=10.1093/glycob/11.1.175;
 RA Hemmerich S., Lee J.-K., Bhakta S., Bistrup A., Ruddle N.R.,
 RA Rosen S.D.,
 RT "Chromosomal localization and genomic organization of the galactose/N-
 RT acetylglucosamine/N-acetylglucosamine 6-O-sulfotransferase gene
 RL family";
 RN Glycobiology 11:75-87(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=2132592; PubMed=11439191; DOI=10.1016/S0092-8674(01)00394-4;
 RA Yen J.-C., Hiraoka N., Pecyniak B., Nakayama J., Elies L.G.,
 RA Rabuka D., Hindegaal O., March J.D., Lowe J.B., Fukuda M.,
 RT "Novel sulfated lymphocyte homing receptors and their control by a
 RL Core1 extension beta 1,3-N-acetylglucosaminyltransferase";
 RN Cell 105:957-969(2001).
 DR EMBL; AF131235; AA033015.1; -
 DR EMBL; AF280088; AA484246.1; -
 DR EMBL; AF149783; AA48417.1; -
 DR GO; GO:0008146; P:sulfotransferase activity; TAS.
 DR GO; GO:0007155; P:cell adhesion; TAS.
 DR GO; GO:0006928; P:cell motility; TAS.
 DR GO; GO:0007267; P:cell-cell signaling; TAS.
 DR GO; GO:0006955; P:immune response; TAS.
 DR GO; GO:0006477; P:protein amino acid sulfation; TAS.
 DR InterPro; IPR000863; Sulfotransferase.
 DR Pfam; PF00685; Sulfotransferase.
 KW Selectin; Transferrase.
 SQ SEQUENCE 386 AA; 45133 MW; OC3BB4022417143A CRC64;

Query Match 49.6%; Score 1019; DB 2; Length 386;
 Best Local Similarity 54.5%; Pred. No. 9.7e-78;
 Matches 216; Conservative 45; Mismatches 107; Indels 28; Gaps 7;
 Qy 14 LLAAQTLLLLPLVSRP-----GPSPAGCEARVHVLLVLSWRSGSFVQGL 59
 Db 2 LLPKMKLLLLVLSQMAILALFFHMYSHNITSSLSMKAPQPMHVLLVLSWRSGSFVQGL 61
 Qy 60 FNOHPDVFYLMPEPAHWVWTTLSQGSAAATLHMAVRLVRSFLCDMDVFDAYL-PWRNLS 118
 Db 62 FGQHPDVFYLMPEPAHWVWMTFKGSTAWMLHMAVRLVRAVFLCDMSVFDAYMEGPRRQS 121
 Qy 119 DLFOAVSRALCSPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 178
 Db 122 SLFOWNSRALCSPACDIIIPQDEIIIPRAHCRLLCSQPPFVVEKACRSYSHVVLKEVRF 181
 Qy 179 FNLQVLYPLSDPALNRIHVLRDPAVRSRQTAALARDNGIVLGTNGTWV-BADP 237
 Db 182 FNLQSLYPLKPSLNHLVLRDPAVRSRRTKGDLMIDSRIVMGQHEQKLKEDQ 241
 Qy 238 GLRVVRECRSHVIAEAATLKP-PPFLRGYRLVREDFLAREPLAEIRALYFTGLSLT 296
 Db 242 PYYVMQVICSQLEIYK--TIQSLPKALQERYLLVRYEDLARAPVAQTSRYEPVGLLEFL 299
 Qy 297 POLBAWHNTHGSGPGARREAFKTSRNALNVQAWRHAPPAKIRRVQBLCAGALQLL 356
 Db 300 PHLTQVHNITRGKMG--DHAFHTNARDALNVQAWRWSLPYEKVSRLQKACGDAMNLL 357
 Qy 357 GYRPVYGEDEQRNALDLVPRGLNGFTWASSTASH 392
 Db 358 GYRHSRSEQRNLLDLLLL-----STWTVPQI 386

RESULT 8
 Q8IV46
 ID Q8IV46 PRELIMINARY; PRT; 370 AA.
 AC Q8IV46;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE CHST4 protein.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

NCBI_TaxID=9606;
 [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.243603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Bhat N.K.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Hsieh F.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalek U., Smailus D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RA Strausberg R.;
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC035282; AAH35282.1; -
 DR GO; GO:0008146; F:sulfotransferase activity; IEA.
 DR InterPro; IPR000863; Sulfotransferase.
 DR Pfam; PF00685; Sulfotransferase_1; 1.
 SQ SEQUENCE 370 AA; 43249 MW; EA8D76EA4E73C625 CRC64;
 Query Match 49.2%; Score 1012; DB 2; Length 370;
 Best Local Similarity 57.9%; Pred. No. 3.6e-77;
 Matches 206; Conservative 43; Mismatches 93; Indels 14; Gaps 6;
 Qy 40 RVHVLVLSWRSGSFVQGLFNOHPDVFYLMPEPAHWVWTTLSQGSAAATLHMAVRLVRSV 99
 Db 26 RVHVLVLSWRSGSFVQGLFGQHPDVFYLMPEPAHWVWMTFKGSTAWMLHMAVRLVRAV 85
 Qy 100 FLCDMDVFDAYL-PWRNLSDLFOAVSRALCSPGAPGAPGAPGAPGAPGAPGAPGAPGAP 158
 Db 86 FLCDMSVFDAYMEGPRRQSLFOWNSRALCSPACDIIIPQDEIIIPRAHCRLLCSQPPF 145
 Qy 159 TLAREACRSYSHVVLKEVRFNQLQVLYPLSDPALNRIHVLRDPAVRSRQTAAL 218
 Db 146 EVVEKACRSYSHVVLKEVRFNQLQSLYPLKPSLNHLVLRDPAVRSRRTKGD 205
 Qy 219 ARNGIVLGTNGTWV-EADPGLRVVRECRSHVIAEAATLKP-PPFLRGYRLVREDFL 276
 Db 206 MIDSRIVMGQHEQKLKEDQPYVVMQVICSQLEIYK--TIQSLPKALQERYLLVRYEDL 263
 Qy 277 AREPLAEIRALYFTGLSLTPQLEAMHNTTHGSGPGARREAFKTSRNALNVQAWRHA 336
 Db 264 ARAPVAQTSRYEPVGLLEFLPHLTQVHNITRGKMG--DHAFHTNARDALNVQAWRWS 321
 Qy 337 LPFAKIRRVQBLCAGALQLLGYRPVYGEDEQRNALDLVPRGLNGFTWASSTASH 392
 Db 322 LPYKVSRLQKACGDAMNLLGYRHSRSEQRNLLDLLLL-----STWTVPQI 370

RESULT 9
 Q9RI11
 ID Q9RI11 PRELIMINARY; PRT; 388 AA.
 AC Q9RI11;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE L-selectin ligand sulfotransferase (Chst4 protein).

RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.,
RT "RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [7]
RP SEQUENCE FROM N.A.
RC STRAIN=CS7BL/6J; TISSUE=Tongue;
RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA Hanagaki T., Hara A., Hiramoto K., Hiraoka T., Hori F.,
RA Inotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
RA Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
RA Okazaki Y., Okido T., Owa C., Saito K., Saito R., Sakai C., Sakai K.,
RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
RA Teima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA Muramatsu M., Hayashizaki Y.,
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF131236; AAD33016.1; -
DR EMBL; AF131236; AAD33016.1; -
DR MGD; MGI:1349479; Chst4.
DR GO; GO:0016021; C:integral to membrane; TAS.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
KW Transferase.
SQ SEQUENCE 388 AA; 44694 MW; D1E9D7796DF8574D CRC64;

Query Match 47.5%; Score 977; DB 2; Length 388;
Best Local Similarity 53.2%; Pred. No. 3.5e-74;
Matches 207; Conservative 43; Mismatches 97; Indels 42; Gaps 7;

QY 14 LLLAQTELLLELVLS-----RPGSSPAGGEARVHVVLSSW 49
DB 2 MLLKKGLLMFLGSQIVVVALFIHMSVHRHLSQRESRP-----VHVVLSSW 50
QY 50 RSGSSFGQLNQHDPVFLMEPAWHVWTTLSQGSAAATLHMAVDLRSVFLCDMDVFDA 109
DB 51 RSGSSFGQLFGQHPDVFYLMPEPAWHVWTTLSQGSAAATLHMAVDLRSVFLCDMDVFDA 110
QY 110 YL-PWRNLSLDFQWAVSRALCSPACAPRGAISSAVCKPCARQSFTLAREACRSY 168
DB 111 YMNPGRKQSLFQWESQSRALCSPACAPRGAISSAVCKPCARQSFTLAREACRSY 170
QY 169 SHVVLKEVRFNQLVPLSDPALNLRIVHLVRDPAVLRSEQTAKALARDNGIVLGT 228
DB 171 GFVVLKEVRFNQLVPLSDPALNLRIVHLVRDPAVLRSEQTAKALARDNGIVLGT 230
QY 229 N-GTWVEADPGLRVVREVCVRSHVRIAE-AFLKPPPPFLGRGYRLVRFDLAREPLAEIRA 286
DB 231 HLETIKEEDQPYAMKICKSQVDIVKAIQTL--PEALQRYLFLRYEDLVRAPLAQTTR 288
QY 287 LYATGLSLTQLEAWIHNTGSGPGARREAFKTSRNALNVSOAWRHALPFAKIRRVQ 346
DB 289 LYKFVGLDFLPHLQTVVYVNRGKMG--QHAFTNARNALNVSOAWRHALPFAKIRRVQ 346
QY 347 ELCAGALLQLGVRPVYSEDEQNLALDLV 375
DB 347 DACGEAMDLLGLVQRSSQEGNSLSDL 375

RESULT 11
Q794G9 PRELIMINARY; PRT; 483 AA.
AC Q794G9;

DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE N-acetylglucosamine-6-O-sulfotransferase (N-Acetylglucosamine 6-O-

DE Sulfotransferase-1).
GN Name=Gn6st-1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Whole embryos;
RX MEDLINE=98380482; PubMed=9712885; DOI=10.1074/jbc.273.35.22577;
RA Uchimura K., Muramatsu H., Kadomatsu K., Fan Q., Kurosawa N.,
RA Mitsuoka C., Kannagi R., Habuchi O., Muramatsu T.,
RT "Molecular cloning and characterization of an N-acetylglucosamine-6-O-
sulfotransferase.";
J. Biol. Chem. 273:22577-22583(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Uchimura K., Kadomatsu K., El-Pasakhany F.M., Singer M.S., Izawa M.,
RA Kannagi R., Takeda N., Rosen S.D., Muramatsu T.,
RT "N-Acetylglucosamine 6-O-Sulfotransferase-1 Regulates Expression of L-
Selectin Ligands and Lymphocyte Homing.";
J. Biol. Chem. 279:35001-35008(2004).
RL EMBL; AB011452; BAA32139.1; -
DR EMBL; AB011452; BAA32139.1; -
DR EMBL; AB011451; BAA32137.1; -
DR GO; GO:0005615; C:extracellular space; TAS.
DR GO; GO:0008146; F:sulfotransferase activity; IDA.
DR GO; GO:0006044; P:N-acetylglucosamine metabolism; TAS.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
KW Transferase.
SQ SEQUENCE 483 AA; 52830 MW; 831FA08F5FEFD70E CRC64;

Query Match 32.2%; Score 663; DB 2; Length 483;
Best Local Similarity 39.5%; Pred. No. 1.8e-47;
Matches 146; Conservative 67; Mismatches 123; Indels 34; Gaps 9;

QY 36 GGEARVHVVLSSWRSGSFGQLFNQHPDVFYLMPEPAWHVWTTLSQGSAAATLHMAVDL 95
DB 113 GGDKRQLVYVFTTWRSGSFGQLFNQHPDVFYLMPEPAWHVWTTLSQGSAAATLHMAVDL 172
QY 96 VRSVFLCDMDVFDAYLP---WRNLSL---FQWAVSRALCSPACAPRGAISS--SEAV 148
DB 173 LSALYRCDLVFQLYSPAGSGRNLTLTGFAATNKKVCCSSPLCPAYRKEVGLVDDRV 232
QY 149 CKPLCARQSFTLAREACRSYSHVVLKEVRFNQLVPLSDPALNLRIVHLVRDPAVL 208
DB 233 CKK-CPQRLARPEECRKYRTVVIKGVRFVDVAVLAPLKDPAIDLVKVLHVRDPAVA 291
QY 209 RSREQTAKALARDN-----GIVLGT--NGTWVEAD--PGLRVVREV 245
DB 292 SSRIRSHGLIRSLQVRSRDPAHRMPPLEAAGHKLGAKEGGMGPADYHALGAMEVI 351
QY 246 CRSHVRIAEAAATLKPPFPFLGRGYRLVRFDLAREPLAEIRALYAFGLSLTPQLEAMTHN 305
DB 352 CNSMAKTLQTA-LQPPDWLQCHVLVRYEDLVGDPVKTLRRVDFVGLLVSPENEQFALN 410
QY 306 ITHGSGPGARREAFKTSRNALNVSOAWRHALPFAKIRRVQELCAGALQLGVRPVYSED 365
DB 411 MTSGSGSSSK--PPVWSARNATQANAWRTALTFOQIKQVEEFCYQPMVAVLGYRVSNSPE 468
QY 366 EORNALDLV 375
DB 469 EVKDLSTLL 478

RESULT 12
O88276 PRELIMINARY; PRT; 530 AA.
ID O88276

AC O88276;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)

DE N-acetylglucosamine-6-O-sulfotransferase long form.
GN Name=Chst2;
OS Mus musculus (Mouse).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Whole embryos;
RX MEDLINE=98380482; PubMed=9712885; DOI=10.1074/jbc.273.35.22577;
RA Uchimura K., Muramatsu H., Kadomatsu K., Fan Q., Kurosawa N.,
RA Mitsuoka C., Kannagi R., Habuchi O., Muramatsu T.;
RT "Molecular cloning and characterization of an N-acetylglucosamine-6-O-
RT sulfotransferase.";
RL J. Biol. Chem. 273:22577-22583 (1998).
DR EMBL; AB011452; BAA32138.1; -.
DR MGD; MGI:1891160; Chst2.
DR GO; GO:0005615; C:extracellular space; TAS.
DR GO; GO:0008146; F:sulfotransferase activity; IDA.
DR GO; GO:0006044; P:N-acetylglucosamine metabolism; TAS.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
KW Transferase.
SQ SEQUENCE 530 AA; 57814 MW; A113E1B735C363EC CRC64;

Query Match 32.2%; Score 663; DB 2; Length 530;
Best Local Similarity 39.5%; Pred. No. 2e-47;
Matches 146; Conservative 67; Mismatches 123; Indels 34; Gaps 9;

Qy 36 GGEARVHLVLSWRSGSFVQLFNQHPDVFYLMPEAWHVWTTLSQGSAAATLHMVRDL 95
Db 160 GGDKRLVYVFTTWRSRSGSFGELEFNQNPVEFFLYEPVHVWQKLYPGDAVSLQGAARDM 219
Qy 96 VRSVFLCDMDVFAYLP---WRNLSDL---FQWAVSRALCSPACSAFPRGAIS--SEAV 148
Db 220 LSALYRCDLSVFLYSPAGSGGRNLTTLGFGAATNKVVCSSPLCPAYRKEVVGLVDDRV 279
Qy 149 CKPLCARQSFLLAREACRSYSHVVLKEVRFNQLVPLSDPALNLRIVHLVRDPAVL 208
Db 280 CKK-CPQRLARFECECRKYRTLVIKGVRFVDVAVLAPLLKDPALDLKVIHLVRDPAVA 338
Qy 209 RSREQTAKALARDN-----GIVLGT--NGTWVAD--PGLRVVREV 245
Db 339 SSRIRSRHGLIRESLQVRSRDPRAHRMPFLEAAGHKLGAKEGMPADYHALGAMEVI 398
Qy 246 CRSHVRIAEATLKPPPLRGYRLVRFEDLAREPLAEIRALYAFGLSLTPGLEAWIHN 305
Db 399 CNSMAKTLQTA-LQPPDWLQGHYLVRYEDLVGDPVKTLRRVYDFVGLLVSPMEQFALN 457
Qy 306 ITHGSGPGARREAFKTSRRNALNVSOAHRHALPFKIRRVQELCAGALQLLGYRVPVSED 365
Db 458 MTSGSGSSSK--PFVVSARNATQAAANWRTALTFOQIKQVEEFCYQPMVILGYRVNSPE 515
Qy 366 EQRNALDLV 375
Db 516 EVKDLSKTL 525

RESULT 13
Q80WV3 PRELIMINARY; PRT; 530 AA.
AC Q80WV3;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Chst2 protein.
OS Mus musculus (Mouse).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Moore T.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strausberg R.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBSJ databases.
DR EMBL; BC051963; AH51963.1; -.
DR MGD; MGI:1891160; Chst2.
DR GO; GO:0005615; C:extracellular space; TAS.
DR GO; GO:0008146; F:sulfotransferase activity; IDA.
DR GO; GO:0006044; P:N-acetylglucosamine metabolism; TAS.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
SQ SEQUENCE 530 AA; 57828 MW; 275363BF15440730 CRC64;

Query Match 32.1%; Score 660; DB 2; Length 530;
Best Local Similarity 39.2%; Pred. No. 3.6e-47;
Matches 145; Conservative 68; Mismatches 123; Indels 34; Gaps 9;

Qy 36 GGEARVHLVLSWRSGSFVQLFNQHPDVFYLMPEAWHVWTTLSQGSAAATLHMVRDL 95
Db 160 GGDKRLVYVFTTWRSRSGSFGELEFNQNPVEFFLYEPVHVWQKLYPGDAVSLQGAARDM 219
Qy 96 VRSVFLCDMDVFAYLP---WRNLSDL---FQWAVSRALCSPACSAFPRGAIS--SEAV 148
Db 220 LSALYRCDLSVFLYSPAGSGGRNLTTLGFGAATNKVVCSSPLCPAYRKEVVGLVDDRV 279
Qy 149 CKPLCARQSFLLAREACRSYSHVVLKEVRFNQLVPLSDPALNLRIVHLVRDPAVL 208
Db 280 CKK-CPQRLARFECECRKYRTLVIKGVRFVDVAVLAPLLKDPALDLKVIHLVRDPAVA 338
Qy 209 RSREQTAKALARDN-----GIVLGT--NGTWVAD--PGLRVVREV 245
Db 339 SSRIRSRHGLIRESLQVRSRDPRAHRMPFLEAAGHKLGAKEGMPADYHALGAMEVI 398
Qy 246 CRSHVRIAEATLKPPPLRGYRLVRFEDLAREPLAEIRALYAFGLSLTPGLEAWIHN 305
Db 399 CNSMAKTLQTA-LQPPDWLQGHYLVRYEDLVGDPVKTLRRVYDFVGLLVSPMEQFALN 457
Qy 306 ITHGSGPGARREAFKTSRRNALNVSOAHRHALPFKIRRVQELCAGALQLLGYRVPVSED 365
Db 458 MTSGSGSSSK--PFVVSARNATQAAANWRTALTFOQIKQVEEFCYQPMVILGYRVNSPE 515
Qy 366 EQRNALDLV 375
Db 516 EVKDLSKTL 525

RESULT 14
Q9UED5 PRELIMINARY; PRT; 483 AA.
ID Q9UED5
AC Q9UED5;

01-MAY-2000 (TrEMBLrel. 13, Created)
01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
N-acetylglucosamine-6-O-sulfotransferase (GlcNAc6ST).
Name=GN6ST;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98391845; PubMed=9722682;
RA Uchimura K., Muramatsu H., Kaname T., Ogawa H., Yamakawa T., Fan Q.,
RA Mitsuoka C., Kannagi R., Habuchi O., Yokoyama I., Yamamura K.,
RA Ozaki T., Nakagawa A., Kadamatsu K., Muramatsu T.,
RT "Human N-acetylglucosamine-6-O-sulfotransferase involved in the
RT biosynthesis of 6-sulfo sialyl Lewis X: molecular cloning, chromosomal
RT mapping, and expression in various organs and tumor cells."; J.
RT J. Biochem. 124:670-678(1998).
[2]
SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=20498786; PubMed=11042394; DOI=10.1016/S0304-4165(00)00136-7;
RA Sakaguchi H., Kitagawa H., Sugahara K.,
RT "Functional expression and genomic structure of human N-
RT acetylglucosamine-6-O-sulfotransferase that transfers sulfate to b-N-
RT acetylglucosamine at the nonreducing end of an N-acetylglucosamine
RT sequence."; Biochim. Biophys. Acta 1523:269-276(2000).
RL Biochim. Biophys. Acta 1523:269-276(2000).
DR EMBL; AB014679; BAA34265.2; -
DR EMBL; AB021124; BAB16886.1; -
DR EMBL; AB021125; BAB16887.1; -
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
KW Transferase.
SQ SEQUENCE 493 AA; 52787 MW; E27797D44931BA18 CRC64;
Query Match 32.0%; Score 657.5; DB 2; Length 483;
Best Local Similarity 38.2%; Pred. No. 5.2e-47;
Matches 147; Conservative 70; Mismatches 125; Indels 43; Gaps 10;
Qy 30 GPSPAG-----GEARVHVLVLSWRSGSFVQGFNQHDPDVFYLMPEPAHVVMTTL 80
Db 98 GVAAPPNGTGTGGVGDKEQLVYVFTTWRSFGSGFFGELFNQNPVEFVFLYEPVHVWVKL 157
Qy 81 SQSAAATLHMVRLVRSVFLCDMDVFDAYLP---WRRNLSDL--FQWAVSRALCSPAC 135
Db 158 YPGDAVSLOGAARDMLSALYRCDLSVQLYSPAGSGGNLTTLGFGAATNKVCSPLC 217
Qy 136 SAFPRGAIS--SEAVCKPLCARQSFLLAREACRSYSHVVLKEVFFNQLVPLLSDPAL 193
Db 218 PAYRKEVVGVLVDDRVCKK-CPQRLARFEBCRYKRTLVKGVRFVDFVAVLAPLRDPAL 276
Qy 194 NLRVHLVDRPRAVLRSEQTAKALARDN-----GIVLGT--NGT 231
Db 277 DLKVIHLVDRPRAVASSRIRSRHGLIRESLOVVRSDPRAHMPPFLEAAGHKLGAKEGV 336
Qy 232 WVEAD-PGLRVVRECVSHVRIAATAATLKPPFLRGYRLVRFFEDLAREPLAEIRALYAF 290
Db 337 GGPADYHALGAMEVICNSMAKTLQTA-LQPPDWLQGHVLYVRYEDLVGDPVKTLLRYVDF 395
Qy 291 TGLSLTPQLEAWIHNTHGSGPGARREAFKTSRNALNVSOAWRHLPFAKIRRVQELCA 350
Db 396 VGLLVSPMEQFALNMTSGSGSSK--PFVVSARNATAQANAWRTALTFTQOIKQVEBFCY 453
Qy 351 GALQLLGYRPVYSEDEQRLALDLV 375
Db 454 QPMVGLGYRVPVSEEVKDLSTLL 478

RESULT 15
QY4C5 PRELIMINARY; PRT; 530 AA.
ID QY4C5; Q9GZNS; Q9Y6F2;
AC QY4C5; Q9GZNS; Q9Y6F2; 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 27, Last annotation update)
DE Long form of N-acetylglucosamine-6-O-sulfotransferase (GlcNAc6ST)
DE Carbohydrate sulfotransferase 2).
DE Name=GN6ST; Synonyms=CHST2;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98391845; PubMed=9722682;
RA Uchimura K., Muramatsu H., Kaname T., Ogawa H., Yamakawa T., Fan Q.,
RA Mitsuoka C., Kannagi R., Habuchi O., Yokoyama I., Yamamura K.,
RA Ozaki T., Nakagawa A., Kadamatsu K., Muramatsu T.,
RT "Human N-acetylglucosamine-6-O-sulfotransferase involved in the
RT biosynthesis of 6-sulfo sialyl Lewis X: molecular cloning, chromosomal
RT mapping, and expression in various organs and tumor cells."; J.
RT J. Biochem. 124:670-678(1998).
[2]
SEQUENCE FROM N.A.
RC TISSUE=Umbilical vein endothelium;
RX MEDLINE=99168906; PubMed=10049591; DOI=10.1006/geno.1998.5653;
RA Li X., Tedder T.F.,
RT "CHST1 and CHST2 sulfotransferases expressed by human vascular
RT endothelial cells: cDNA cloning, expression, and chromosomal
RT localization."; Genomics 55:345-347(1999).
RL Genomics 55:345-347(1999).
DR EMBL; AB014680; BAA34266.2; -
DR EMBL; AF083066; AAD20981.1; -
DR Genew; HGNC:1970; CHST2.
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase..
DR Pfam; PF00685; Sulfotransfer_1; 1.
KW Transferase.
SQ SEQUENCE 530 AA; 57857 MW; A82CA22789D5651B CRC64;
Query Match 32.0%; Score 657.5; DB 2; Length 530;
Best Local Similarity 38.2%; Pred. No. 5.9e-47;
Matches 147; Conservative 70; Mismatches 125; Indels 43; Gaps 10;
Qy 30 GPSPAG-----GEARVHVLVLSWRSGSFVQGFNQHDPDVFYLMPEPAHVVMTTL 80
Db 145 GVAAPPNGTGTGGVGDKEQLVYVFTTWRSFGSGFFGELFNQNPVEFVFLYEPVHVWVKL 204
Qy 81 SQSAAATLHMVRLVRSVFLCDMDVFDAYLP---WRRNLSDL--FQWAVSRALCSPAC 135
Db 205 YPGDAVSLOGAARDMLSALYRCDLSVQLYSPAGSGGNLTTLGFGAATNKVCSPLC 264
Qy 136 SAFPRGAIS--SEAVCKPLCARQSFLLAREACRSYSHVVLKEVFFNQLVPLLSDPAL 193
Db 265 PAYRKEVVGVLVDDRVCKK-CPQRLARFEBCRYKRTLVKGVRFVDFVAVLAPLRDPAL 323
Qy 194 NLRVHLVDRPRAVLRSEQTAKALARDN-----GIVLGT--NGT 231
Db 324 DLKVIHLVDRPRAVASSRIRSRHGLIRESLOVVRSDPRAHMPPFLEAAGHKLGAKEGV 383
Qy 232 WVEAD-PGLRVVRECVSHVRIAATAATLKPPFLRGYRLVRFFEDLAREPLAEIRALYAF 290
Db 384 GGPADYHALGAMEVICNSMAKTLQTA-LQPPDWLQGHVLYVRYEDLVGDPVKTLLRYVDF 442
Qy 291 TGLSLTPQLEAWIHNTHGSGPGARREAFKTSRNALNVSOAWRHLPFAKIRRVQELCA 350
Db 443 VGLLVSPMEQFALNMTSGSGSSK--PFVVSARNATAQANAWRTALTFTQOIKQVEBFCY 500
Qy 351 GALQLLGYRPVYSEDEQRLALDLV 375

Db : : : : | : | : : | : : : | :
501 QPMVGLGYERVNSPEEVKDL\$KTL\$ 525

Search completed: June 23, 2005, 08:49:01
Job time : 44.3472 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2005, 08:32:43 ; Search time 132.064 Seconds
(without alignments)
3578.727 Million cell updates/sec

Title: US-10-697-828-15

Perfect score: 6532

Sequence: 1 MPRGGAPPIMALMFTGHL.....LIENICWTLMDRLGYPKFMD 1222

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

A_Geneseq_16Dec04:*

- 1: Geneseqp1980s:*
- 2: Geneseqp1990s:*
- 3: Geneseqp2000s:*
- 4: Geneseqp2001s:*
- 5: Geneseqp2002s:*
- 6: Geneseqp2003as:*
- 7: Geneseqp2003bs:*
- 8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|---------------------|
| 1 | 6532 | 100.0 | 1222 | 4 | AAV72642 Human gly |
| 2 | 6509 | 99.6 | 1222 | 6 | ABU11849 Human sec |
| 3 | 6465 | 99.0 | 1212 | 6 | AAE33541 Human nov |
| 4 | 5899.5 | 90.3 | 1207 | 4 | AAV72643 Mouse gly |
| 5 | 5899.5 | 90.3 | 1207 | 6 | AAE33542 Human nov |
| 6 | 4037 | 61.8 | 755 | 4 | AAAB93735 Human pro |
| 7 | 4037 | 61.8 | 755 | 8 | ADP56673 Human pro |
| 8 | 3147 | 48.2 | 596 | 4 | AAV72641 Human gly |
| 9 | 2597 | 39.8 | 480 | 4 | AAAB94410 Human pro |
| 10 | 1753.5 | 26.8 | 958 | 3 | AAAY51120 Human SAR |
| 11 | 1753.5 | 26.8 | 958 | 8 | ADL83102 Human PRO |
| 12 | 1331 | 20.4 | 314 | 4 | AAAB93826 Human pro |
| 13 | 898 | 13.7 | 625 | 8 | ADR10287 Human pro |
| 14 | 637 | 9.8 | 128 | 4 | ADG27796 Human nov |
| 15 | 568 | 8.7 | 125 | 8 | ADJ12217 Human sec |
| 16 | 568 | 8.7 | 126 | 2 | AAAY14415 Human sec |
| 17 | 431 | 6.6 | 456 | 8 | ADP30411 Human sec |
| 18 | 368 | 5.6 | 218 | 4 | AAAM24176 Human EST |
| 19 | 348.5 | 5.3 | 474 | 3 | AAAB34722 Human sec |
| 20 | 147 | 2.3 | 410 | 7 | ADN95522 Human BEC |
| 21 | 147 | 2.3 | 411 | 2 | AAAM61100 Keratan s |
| 22 | 147 | 2.3 | 411 | 5 | AAE25356 Human cho |
| 23 | 147 | 2.3 | 411 | 6 | ABU03503 Angiogene |
| 24 | 147 | 2.3 | 411 | 7 | ADJ68589 Human hea |
| 25 | 147 | 2.3 | 411 | 8 | ADQ18725 Human sof |

| | | | | | | |
|----|-----|-----|------|---|-----------|---------------------|
| 26 | 145 | 2.2 | 2504 | 2 | AAW17871 | AAW17871 Photorhab |
| 27 | 145 | 2.2 | 2504 | 2 | AAW56557 | AAW56557 Toxin Tcb |
| 28 | 145 | 2.2 | 2504 | 4 | AAAB72610 | AAAB72610 Photorhab |
| 29 | 145 | 2.2 | 2504 | 5 | ABG32654 | ABG32654 P. lumine |
| 30 | 145 | 2.2 | 2504 | 8 | ADR21539 | ADR21539 Photorhab |
| 31 | 145 | 2.2 | 2505 | 4 | AAAB72612 | AAAB72612 Modified |
| 32 | 135 | 2.1 | 1844 | 2 | AAW18302 | AAW18302 Photorhab |
| 33 | 135 | 2.1 | 1844 | 2 | AAW56558 | AAW56558 Toxin Tcb |
| 34 | 134 | 2.1 | 483 | 2 | AAV31657 | AAV31657 Mouse N-a |
| 35 | 133 | 2.0 | 484 | 2 | AAV31657 | AAV31657 Human N-a |
| 36 | 133 | 2.0 | 530 | 4 | AAAB95367 | AAAB95367 Human pro |
| 37 | 133 | 2.0 | 530 | 8 | ADQ18590 | ADQ18590 Human sof |
| 38 | 133 | 2.0 | 531 | 5 | AAU69414 | AAU69414 Lung smal |
| 39 | 133 | 2.0 | 531 | 8 | ADRI14283 | ADRI14283 Human NF- |
| 40 | 130 | 2.0 | 1456 | 2 | AAAR24033 | AAAR24033 Soluble m |
| 41 | 128 | 2.0 | 183 | 4 | ABAB68582 | ABAB68582 Drosophil |
| 42 | 127 | 1.9 | 1456 | 6 | ABU79121 | ABU79121 Angiostat |
| 43 | 127 | 1.9 | 1456 | 7 | ADFA33375 | ADFA33375 Mannose r |
| 44 | 127 | 1.9 | 1456 | 7 | ADN39899 | ADN39899 Cancer/an |
| 45 | 127 | 1.9 | 1456 | 7 | ADN95147 | ADN95147 Human BEC |

ALIGNMENTS

RESULT 1

AAV72642

ID AAV72642 standard; protein; 1222 AA.

XX AAV72642;

XX 02-MAY-2001 (first entry)

XX Human glycosyl sulfotransferase-6 (GST-6).

XX Human; glycosyl sulfotransferase-6; GST-6; immunosuppressive; therapy;
selectin binding inhibitor; gene therapy; inflammation;
systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenailitis;
Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
asthma; hypersensitivity; rheumatic fever; tissue rejection.

OS Homo sapiens.

XX Key Location/Qualifiers

Domain 851..1223
/label= C-terminal_sulfotransferase_domain

XX WO200106015-A1.

XX 25-JAN-2001.

XX 19-JUL-2000; 2000WO-US019741.

XX 20-JUL-1999; 99US-0144694P.

XX 13-JUN-2000; 2000US-00593828.

XX (REGC) UNIV CALIFORNIA.

XX Rosen SD, Lee JK, Hemmerich S;

XX WPI; 2001-138471/14.

XX N-PSDB; AAD02702, AAD02704.

XX New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
diagnostic and therapeutic agent screening applications.

XX Claim 3; Fig 5B; 128pp; English.

XX The present sequence is human glycosyl sulfotransferase-6 (GST-6). GST is

a type 2 membrane protein useful for inhibiting a binding event between a selectin and a selectin ligand, which comprises contacting the selectin with a non-sulphated selectin ligand, GST and a small molecular agent that inhibits the sulphation activity of GST. GST is also useful in inhibiting a selectin mediated binding event. GST is useful in gene therapy to treat disorders such as acute or chronic inflammation, CC systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes, CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis, pericarditis, myocardiitis, regional enteritis, adult respiratory distress CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis, CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection CC during transplantation

XX SQ Sequence 1222 AA;

Query Match 100.0%; Score 6532; DB 4; Length 1222;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1222; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

| | | | |
|----|-----|--|-----|
| Qy | 1 | MPKGGAPPIMALMFTGHLFLALLMFAFSTPEESVSNYSENAVFTDDIDQFKTKVQDF | 60 |
| Db | 1 | MPKGGAPPIMALMFTGHLFLALLMFAFSTPEESVSNYSENAVFTDDIDQFKTKVQDF | 60 |
| Qy | 61 | RPNQKLKSMHPSLYFDAGEIOAMRQKSRASHLHLFRAIRSAVTVMLSNPYYLPPPKH | 120 |
| Db | 61 | RPNQKLKSMHPSLYFDAGEIOAMRQKSRASHLHLFRAIRSAVTVMLSNPYYLPPPKH | 120 |
| Qy | 121 | ADFAAKWNEIYGNLPPALYCLCPEDKVAPEFVLEYMWRMVGKDWLVENAPGDEVPI | 180 |
| Db | 121 | ADFAAKWNEIYGNLPPALYCLCPEDKVAPEFVLEYMWRMVGKDWLVENAPGDEVPI | 180 |
| Qy | 181 | GHSLTGFATAFDLYNLLDNHRRQKYLEKIWIITEEMYEYSKVRSGKQLLNHHQATNMI | 240 |
| Db | 181 | GHSLTGFATAFDLYNLLDNHRRQKYLEKIWIITEEMYEYSKVRSGKQLLNHHQATNMI | 240 |
| Qy | 241 | ALLTGALVTGVDGSKANIWKQAVDVMEKTMFLNHHIVDGLDEGVAIGSVTAKSVTQY | 300 |
| Db | 241 | ALLTGALVTGVDGSKANIWKQAVDVMEKTMFLNHHIVDGLDEGVAIGSVTAKSVTQY | 300 |
| Qy | 301 | VFLAQHRFNINLLDNWKKHFWFYATLLPGFQRTVGIADSNYNWFYSPESQLVFLDKF | 360 |
| Db | 301 | VFLAQHRFNINLLDNWKKHFWFYATLLPGFQRTVGIADSNYNWFYSPESQLVFLDKF | 360 |
| Qy | 361 | ILKXGAGNWLAAQIRKXRPKDGMPVFPSTAQRWSTLHTEIWDYDQPTPPADYGTAKTH | 420 |
| Db | 361 | ILKXGAGNWLAAQIRKXRPKDGMPVFPSTAQRWSTLHTEIWDYDQPTPPADYGTAKTH | 420 |
| Qy | 421 | TFPNWGVVTVGAGLPNTQNTVTFYFKSGKLGRNAVYDVIHFQPSYWDGWRSPNGHEHP | 480 |
| Db | 421 | TFPNWGVVTVGAGLPNTQNTVTFYFKSGKLGRNAVYDVIHFQPSYWDGWRSPNGHEHP | 480 |
| Qy | 481 | DQNSFTFAPNGQVVFSEALYGPKLSHLNNVLPAPSPQCNKPWEGQGECAQMLKWTG | 540 |
| Db | 481 | DQNSFTFAPNGQVVFSEALYGPKLSHLNNVLPAPSPQCNKPWEGQGECAQMLKWTG | 540 |
| Qy | 541 | EEVGDAAGEIITASQHGEMVVFVSGEASVAYSSAMRLKSVYRALLLNSQTLVVDHIERQ | 600 |
| Db | 541 | EEVGDAAGEIITASQHGEMVVFVSGEASVAYSSAMRLKSVYRALLLNSQTLVVDHIERQ | 600 |
| Qy | 601 | EDSPINSVSAFFNLIDIFKYIIPKFNRYNGAMQVDAHYKQMFWDHGHGNSPMASIOE | 660 |
| Db | 601 | EDSPINSVSAFFNLIDIFKYIIPKFNRYNGAMQVDAHYKQMFWDHGHGNSPMASIOE | 660 |
| Qy | 661 | AEQAAEFKKRWTOFVNVTFQMESTIIRIAVVFYGPYINVSRCRFDISSNPGIQLISLVNN | 720 |
| Db | 661 | AEQAAEFKKRWTOFVNVTFQMESTIIRIAVVFYGPYINVSRCRFDISSNPGIQLISLVNN | 720 |
| Qy | 721 | TEHVVISVTDYHNLKTRFNLYLGGFGFASVADQGOITRFGLTQAIKVPVRHDIRIIPFGF | 780 |
| Db | 721 | TEHVVISVTDYHNLKTRFNLYLGGFGFASVADQGOITRFGLTQAIKVPVRHDIRIIPFGF | 780 |

| | | | |
|----------|---|---|------|
| Qy | 781 | KFNIAVGILICISLVILTFQWRFYLSFRKLMRWILILVIALWFIELLDVWSTCSOPICAK | 840 |
| Db | 781 | KFNIAVGILICISLVILTFQWRFYLSFRKLMRWILILVIALWFIELLDVWSTCSOPICAK | 840 |
| Qy | 841 | WTRTEAGSKLSLSEGHMDLPDVVITSLPGSGAEILKQLFFNSDDFLYIRVPTAYIDI | 900 |
| Db | 841 | WTRTEAGSKLSLSEGHMDLPDVVITSLPGSGAEILKQLFFNSDDFLYIRVPTAYIDI | 900 |
| Qy | 901 | PETELEIDSFVDACSWKVSIDIRSGHFRLLRGWLSQVODTKLHLQNIHLHHPNRKLAQY | 960 |
| Db | 901 | PETELEIDSFVDACSWKVSIDIRSGHFRLLRGWLSQVODTKLHLQNIHLHHPNRKLAQY | 960 |
| Qy | 961 | FAMNKDKRKKRRESLPEQBSQMKGAFDRDAEYIRALRRHLVYYPSPARPVLSLSSGSGWT | 1020 |
| Db | 961 | FAMNKDKRKKRRESLPEQBSQMKGAFDRDAEYIRALRRHLVYYPSPARPVLSLSSGSGWT | 1020 |
| Qy | 1021 | LKLHFFQEVGLGASMRALYIVRDPRAWIYSLMYNSKPSLYSLKNVPEHLAKLFKIEGGKKG | 1080 |
| Db | 1021 | LKLHFFQEVGLGASMRALYIVRDPRAWIYSLMYNSKPSLYSLKNVPEHLAKLFKIEGGKKG | 1080 |
| Qy | 1081 | CNLSNGYAFEPYELRKELSKSNVSLLSHLWLAANTAAALRINTDILLPTSYQLVKFEDI | 1140 |
| Db | 1081 | CNLSNGYAFEPYELRKELSKSNVSLLSHLWLAANTAAALRINTDILLPTSYQLVKFEDI | 1140 |
| Qy | 1141 | VHFPOKTTTERIFAFIIGPLSPASLNQILFATSTNLFYLPYEGEISPTNTNVKQNLPRDE | 1200 |
| Db | 1141 | VHFPOKTTTERIFAFIIGPLSPASLNQILFATSTNLFYLPYEGEISPTNTNVKQNLPRDE | 1200 |
| Qy | 1201 | IKLIENICWTLMMDRLGYPKFMD 1222 | |
| Db | 1201 | IKLIENICWTLMMDRLGYPKFMD 1222 | |
| RESULT 2 | | | |
| ABU11849 | | | |
| ID | ABU11849 | standard; protein; 1222 AA. | |
| XX | ABU11849; | | |
| AC | ABU11849; | | |
| XX | | | |
| DT | 12-FEB-2003 | (first entry) | |
| XX | | | |
| DE | Human secreted protein SECP-4, | INCYTE 3441255CD1. | |
| KW | Human; SECP; secreted protein; micro-array; liver disease; hepatitis; cirrhosis; cell proliferative disease; cancer; atherosclerosis; | | |
| KW | neurological disorder; epilepsy; Huntington's disease; stroke; | | |
| KW | cardiovascular disorder; hypertension; angina pectoris; allergy; | | |
| KW | myocardial infarction; immune disorder; inflammatory disorder; AIDS; | | |
| KW | hypothyroidism; acquired immunodeficiency syndrome; Cushing's syndrome; developmental disorder. | | |
| OS | Homo sapiens. | | |
| XX | | | |
| PN | WO200286069-A2. | | |
| XX | | | |
| PD | 31-OCT-2002. | | |
| XX | | | |
| PF | 19-APR-2002; 2002WO-US012464. | | |
| XX | | | |
| PR | 20-APR-2001; 2001US-0285207P. | | |
| PR | 27-APR-2001; 2001US-0287114P. | | |
| PR | 03-MAY-2001; 2001US-0288640P. | | |
| PR | 11-MAY-2001; 2001US-0290516P. | | |
| PR | 18-MAY-2001; 2001US-0292184P. | | |
| PR | 21-DEC-2001; 2001US-0343553P. | | |
| PR | 13-FEB-2002; 2002US-0357002P. | | |
| PR | 20-FEB-2002; 2002US-0358279P. | | |
| PR | 19-MAR-2002; 2002US-0366041P. | | |
| XX | | | |
| PA | (INCY-) INCYTE GENOMICS INC. | | |
| XX | | | |
| PI | Klammer AA, Hafalia AJA, Duggan BM, Warren BA, Emerling BM; | | |

PI Tribouley CM, Arvizu CS, Honchell CD, Nguyen DB, Kallick DA, Yue H;
 PI Au-Young JK, Ramkumar J, Li JX, Thangavelu K, Gietzen KJ, Ding L;
 PI Baughn MR, Yao MG, Walia NK, Mason PM, Lal PG, Graul RC, Reddy R;
 PI Becha SD, Sapperstein SK, Richardson TW, Tran UK, Elliott VS;
 PI Tang YT, Azimzai Y, Yan L, Xu Y;
 XX WPI; 2003-093118/08.
 DR N-PSDB; ABX49951.
 XX
 XX
 PT New human secreted proteins (SECP) useful for diagnosing, treating and
 PT preventing diseases or conditions associated with the aberrant SECP
 PT expression e.g. cancer, AIDS, atherosclerosis, epilepsy, allergies,
 PT hepatitis, cirrhosis.
 XX
 XX
 PS Claim 1: Page 137-140; 192pp; English.
 XX
 CC The invention relates to an isolated polypeptide comprising any of 30
 CC secreted human proteins (SECP1-SECP30) appearing as ABU1846-ABU1875, a
 CC naturally occurring amino acid sequence at least 90-98 % identical to the
 CC sequences, or a biologically active or immunogenic fragment of the
 CC polypeptide. Also included are an isolated polynucleotide encoding SECP,
 CC (including a polynucleotide sequence at least 90-98 % identical to the
 CC sequences, their complements, RNA equivalents or fragments comprising at
 CC least 60 contiguous nucleotides) a recombinant polynucleotide comprising
 CC a promoter sequence operably linked to the SECP polynucleotide, a cell
 CC comprising the recombinant polynucleotide, a transgenic organism
 CC comprising the recombinant polynucleotide, an anti-SECP antibody,
 CC screening for ant/agonists of SECP, generating an expression profile of a
 CC sample containing the polynucleotides and an array comprising different
 CC nucleotide molecules affixed at distinct physical locations on a solid
 CC substrate, where at least one nucleotide molecule comprises a first
 CC oligonucleotide or polynucleotide sequence specifically hybridisable with
 CC at least 30 contiguous nucleotides of the target polynucleotide. The
 CC polypeptides and polynucleotides are useful in diagnosing, treating and
 CC preventing diseases or conditions associated with the decreased
 CC expression or overexpression of SECP, such as liver (e.g. hepatitis,
 CC cirrhosis), cell proliferative (e.g. cancer, atherosclerosis),
 CC neurological (e.g. epilepsy, Huntington's disease, stroke),
 CC cardiovascular (e.g. hypertension, angina pectoris, myocardial
 CC infarction), immune/inflammatory (e.g. acquired immunodeficiency syndrome
 CC (AIDS), allergies) and developmental (e.g. Hypothyroidism, Cushing's
 CC syndrome) disorders (many other diseases and conditions are given in the
 CC specification). These are also useful in assessing the effects of
 CC exogenous compounds on the expression of nucleic acid and amino acid
 CC sequences of SECP. The SECP or its fragments are useful in screening
 CC compounds for effectiveness as agonist or antagonist of the polypeptides,
 CC or in altering the expression of the target polynucleotide and compounds
 CC that specifically bind to or modulate the activity of the polypeptide.
 CC The micro-array is useful in monitoring or measuring protein-protein
 CC interactions, drug-target interactions, and gene expression profiles. The
 CC present sequence is a SECP protein of the invention
 XX
 SQ Sequence 1222 AA;

Query Match 99.6%; Score 6509; DB 6; Length 1222;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 1219; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 Qy 1 MPKGGAPPWIMALMFTGHLFLALLMFASTFESSVSNYSSEWAVFTDDIDQFKTKQVDF 60
 Db 1 MPKGGAPPWIMALMFTGHLFLALLMFASTFESSVSNYSSEWAVFTDDIDQFKTKQVDF 60
 Qy 61 RPNOKLKSMLHPSLYFDAGIOAMRQKSRASHLHLFRAITSATVMSLNTYYLPPPKH 120
 Db 61 RPNOKLKSMLHPSLYFDAGIOAMRQKSRASHLHLFRAITSATVMSLNTYYLPPPKH 120
 Qy 121 ADFAAKWNEIYGNLPLALYCLLPEDKVAFFVLEYMDRMVGYKDWLVENAPGDEVPI 180
 Db 121 ADFAAKWNEIYGNLPLALYCLLPEDKVAFFVLEYMDRMVGYKDWLVENAPGDEVPI 180
 Qy 181 GHSLTGTAFADFLYNLDNHRROKYLEKIWIITEEMVEYSKVRSGWQLLHNHQAATMI 240
 Db 181 VHSLTGTAFADFLYNLDNHRROKYLEKIWIITEEMVEYSKVRSGWQLLHNHQAATMI 240

Qy 241 ALLTGALVTGVDKSGKANIKQAVDVMKTMFLNHHIVDGLSDEGVAYGYSYAKSVTQY 300
 Db 241 ALLTGALVTGVDKSGKANIKQAVDVMKTMFLNHHIVDGLSDEGVAYGYSYAKSVTQY 300
 Qy 301 VFLAQHFNNINNDNNWLVKMHFWFYATLLPGFQRTVGIADSNYNWYFSGESQVFLDKF 360
 Db 301 VFLAQHFNNINNDNNWLVKMHFWFYATLLPGFQRTVGIADSNYNWYFSGESQVFLDKF 360
 Qy 361 ILKNGAGNMLAQOIRKHRPKDGMVSTAQRMSTLHTEYIWDYDQTPQPPADYGTAKIH 420
 Db 361 ILKNGAGNMLAQOIRKHRPKDGMVSTAQRMSTLHTEYIWDYDQTPQPPADYGTAKIH 420
 Qy 421 TFPNWGVVITYGAGLPNTQNTTTFVSKSGKLGRAVVDIVHFQPYSWIDGWRSPNPGHEHP 480
 Db 421 TFPNWGVVITYGAGLPNTQNTTTFVSKSGKLGRAVVDIVHFQPYSWIDGWRSPNPGHEHP 480
 Qy 481 DQNSFTFAPNGQVVFSEALYGPKLHLNNVLFVAPSPSSQCNKPWEGQLGECQWLKWTG 540
 Db 481 DQNSFTFAPNGQVVFSEALYGPKLHLNNVLFVAPSPSSQCNKPWEGQLGECQWLKWTG 540
 Qy 541 EEVGDAAGEIITASOHEGMVFSGEAVSSAMRLKSVYRALLLLNSQTLVVVDHIERQ 600
 Db 541 EEVGDAAGEIITASOHEGMVFSGEAVSSAMRLKSVYRALLLLNSQTLVVVDHIERQ 600
 Qy 601 EDSPINSVSFAFFHNLDIDFKYIYPKPMNRNGAMDMVDAYHKMFDFDHGNSPMASIQE 660
 Db 601 EDSPINSVSFAFFHNLDIDFKYIYPKPMNRNGAMDMVDAYHKMFDFDHGNSPMASIQE 660
 Qy 661 AEQAAPFKKWTQFVNVTFQMESTITRIAYVFGPYINVSCTPIDSNSPGLQISLVNNV 720
 Db 661 AEQAAPFKKWTQFVNVTFQMESTITRIAYVFGPYINVSCTPIDSNSPGLQISLVNNV 720
 Qy 721 TEHVSVITVDYHNLKTRFNVLGGGFASVADQOITRFGLGTOAIVKPVHRDRIIFPGF 780
 Db 721 TEHVSVITVDYHNLKTRFNVLGGGFASVADQOITRFGLGTOAIVKPVHRDRIIFPGF 780
 Qy 781 KFNIAVGLIICISIVILTFQWRVFLSPKLMRWLILVIALWFIELDDVMVSTCSQPICAK 840
 Db 781 KFNIAVGLIICISIVILTFQWRVFLSPKLMRWLILVIALWFIELDDVMVSTCSQPICAK 840
 Qy 841 WTRTEAGSKKSLSSSEGHMDLPDVTITSLPGSGAEILKOLFNNSSDFLIRVPTAYIDI 900
 Db 841 WTRTEAGSKKSLSSSEGHMDLPDVTITSLPGSGAEILKOLFNNSSDFLIRVPTAYIDI 900
 Qy 901 PETELEIDSFVDACEWKVSDIRSGHFRLLRGLQSLVQDTKLHLQNLHLHPEPNRGKLAQY 960
 Db 901 PETELEIDSFVDACEWKVSDIRSGHFRLLRGLQSLVQDTKLHLQNLHLHPEPNRGKLAQY 960
 Qy 961 FANMKDKKRFKRRRESLPEQORSQMKGAFDRDAEYIRALRRHLVYYPARPVLSSSGSWT 1020
 Db 961 FANMKDKKRFKRRRESLPEQORSQMKGAFDRDAEYIRALRRHLVYYPARPVLSSSGSWT 1020
 Qy 1021 LKLFHFQEVILGASNRALYIVRDPRAWIYMLYNSKPSLSYLNKVPHEHLAKLFKIEGGK 1080
 Db 1021 LKLFHFQEVILGASNRALYIVRDPRAWIYMLYNSKPSLSYLNKVPHEHLAKLFKIEGGK 1080
 Qy 1081 CNLNSGYAPFEPYELRKELSKSNVSLSHLWLANTAARINTDLPYSLVQVLPKEDI 1140
 Db 1081 CNLNSGYAPFEPYELRKELSKSNVSLSHLWLANTAARINTDLPYSLVQVLPKEDI 1140
 Qy 1141 VHPQKTERIFAPLGIPLSPASINQILFATSTNLFLPYLPEGEISPTNTNVNKNLPRDE 1200
 Db 1141 VHPQKTERIFAPLGIPLSPASINQILFATSTNLFLPYLPEGEISPTNTNVNKNLPRDE 1200
 Qy 1201 IKLIENICWTMDRLGYPKFM 1222
 Db 1201 IKLIENICWTMDRLGYPKFM 1222
 RESULT 3
 AAE33541
 ID AAE33541 standard; protein; 1212 AA.

| | | |
|----|--|--|
| XX | AAE33541; | |
| AC | 16-APR-2003 (first entry) | |
| DT | Human novel CpG-associated gene 1 (NCAG1) encoded protein #1. | |
| DE | Human; novel CpG-associated gene 1; bipolar disorder; neuroprotective; | |
| DE | NCAG1; mood disorder; chromosome 18. | |
| XX | Homo sapiens. | |
| KW | WO2002101044-A2. | |
| XX | 19-DEC-2002. | |
| XX | 06-JUN-2002; 2002WO-EP006316. | |
| XX | 11-JUN-2001; 2001EP-00202214. | |
| XX | (JANC) JANSSEN PHARM NV. | |
| PA | Del-Favero JPL, Van Broeckhoven C; | |
| PI | WPI; 2003-148807/14. | |
| DR | N-PSDB; AAD50032. | |
| XX | New brain expressed genes (designated novel CpG-associated Gene 1 | |
| PT | (NCAG1)) and its encoded protein, useful as diagnostic markers for | |
| PT | bipolar or mood disorders, and as targets for developing drugs for the | |
| PT | treatment these disorders. | |
| XX | Claim 14; Col 38-42; 26pp; English. | |
| CC | The invention relates to novel CpG-associated gene 1 (NCAG1) brain- | |
| CC | expressed gene and its encoded protein. The NCAG1 nucleic acid or its | |
| CC | encoded protein is useful as a diagnostic marker for bipolar disorder | |
| CC | such as mood disorders. They are also useful as targets for developing | |
| CC | drugs, as well as for target validation, for the treatment of bipolar | |
| CC | disorders. The present sequence is human NCAG1 protein. NCAG1 gene is | |
| CC | located at chromosome 18 | |
| XX | Sequence 1212 AA; | |
| SQ | Query Match 99.0%; Score 6465; DB 6; Length 1212; | |
| | Best Local Similarity 99.9%; Pred. No. 0; | |
| | Matches 1211; Conservative 0; Mismatches 1; Indels 0; Gaps 0; | |
| Qy | 11 MALMFTGHLHFLALLMPAFSTFEESVSNYSEWAVFTDDIDQFKTKVQVDFRPNQKLKXSM 70 | |
| Db | 1 MALMFTGHLHFLALLMPAFSTFEESVSNYSEWAVFTDDIDQFKTKVQVDFRPNQKLKXSM 60 | |
| Qy | 71 LHPSLYFDAGEIQAMRQKSRASHLHLFRAIRSAVTVMLSNPTYYLPPPKHADPAKAKWNEI 130 | |
| Db | 61 LHPSLYFDAGEIQAMRQKSRASHLHLFRAIRSAVTVMLSNPTYYLPPPKHADPAKAKWNEI 120 | |
| Qy | 131 YGNLPLALYCLLPEDKVAFFEVLEYMDRMVGYKDWLVENAPGDEVPIGHSLTGFATA 190 | |
| Db | 121 YGNLPLALYCLLPEDKVAFFEVLEYMDRMVGYKDWLVENAPGDEVPIGHSLTGFATA 180 | |
| Qy | 191 FDFLYNLLDNHRKQKYLEKTLWVTEEMYEYSKVRSGKOLLHNHQAATNMIALLTGALVTG 250 | |
| Db | 181 FDFLYNLLDNHRKQKYLEKTLWVTEEMYEYSKVRSGKOLLHNHQAATNMIALLTGALVTG 240 | |
| Qy | 251 VDKGSKANIWKQAVVDVMEKTMFLNHIVDGSLDEGVAYGYSYAKSVTVQVFLAQRHFN 310 | |
| Db | 241 VDKGSKANIWKQAVVDVMEKTMFLNHIVDGSLDEGVAYGYSYAKSVTVQVFLAQRHFN 300 | |
| Qy | 311 NNLDDNWLKMHFWFYATLLPGQRTVGIADSNYNHFGPESOLVFLDKFKLKGAGNWL 370 | |
| Db | 301 NNLDDNWLKMHFWFYATLLPGQRTVGIADSNYNHFGPESOLVFLDKFKLKGAGNWL 360 | |
| Qy | 371 AQQIRKHRPKDGPMPVSTAQRWSTLHTEYIWYDPQLTPQPPADYGTAKIHTFPNGWVTV 430 | |

| | | |
|----------|---|--|
| Db | 361 AQQIRKHRPKDGPMPVSTAQRWSTLHTEYIWYDPQLTPQPPADYGTAKIHTFPNGWVTV 420 | |
| Qy | 431 GAGLENTQTNTFVSKGKLGGRAYDIVHFQPYSWIDGWRSFNFGCHEHPDQNSFTFAFN 490 | |
| Db | 421 GAGLENTQTNTFVSKGKLGGRAYDIVHFQPYSWIDGWRSFNFGCHEHPDQNSFTFAFN 480 | |
| Qy | 491 GQVFVSEALYGPKLSHLNNLVFAPSPSSQCNKPWEGOLGCEAOMLKWGTGEEVGAAGBI 550 | |
| Db | 481 GQVFVSEALYGPKLSHLNNLVFAPSPSSQCNKPWEGOLGCEAOMLKWGTGEEVGAAGBI 540 | |
| Qy | 551 ITASQHGEMVFSGEAVSAYSSAMRLKSVYRALLLNLSQTLVVVDHIERQEDSPINSVA 610 | |
| Db | 541 ITASQHGEMVFSGEAVSAYSSAMRLKSVYRALLLNLSQTLVVVDHIERQEDSPINSVA 600 | |
| Qy | 611 PFPHNLDIDFKYIPYKPMRNYNGAMDMVDADAHYKMFWDHGHGNSPMASIOEAQAAEFKCR 670 | |
| Db | 601 PFPHNLDIDFKYIPYKPMRNYNGAMDMVDADAHYKMFWDHGHGNSPMASIOEAQAAEFKCR 660 | |
| Qy | 671 WTQFVNVTFOBESTITRIAYVYFGPYINVSSCRFTDSSNPGLOISLVNNTNTEHVVSIVTD 730 | |
| Db | 661 WTQFVNVTFOBESTITRIAYVYFGPYINVSSCRFTDSSNPGLOISLVNNTNTEHVVSIVTD 720 | |
| Qy | 731 YHNLKTRFNLYLGGFGFASVADQGITRFGLGTAIVKPVVRHDDRIIFPFGFKENIAVGLIL 790 | |
| Db | 721 YHNLKTRFNLYLGGFGFASVADQGITRFGLGTAIVKPVVRHDDRIIFPFGFKENIAVGLIL 780 | |
| Qy | 791 CISLVILTFQWRFYLSFRKLARWILILVIALWFIELLDVWSTCSQPICAKWTTRTEAGSK 850 | |
| Db | 781 CISLVILTFQWRFYLSFRKLARWILILVIALWFIELLDVWSTCSQPICAKWTTRTEAGSK 840 | |
| Qy | 851 KSLSEGHMDLPDVVITSLPGSGAEILKOLFNFSSDFLYIRVPTAYIDIPETELEIDSF 910 | |
| Db | 841 KSLSEGHMDLPDVVITSLPGSGAEILKOLFNFSSDFLYIRVPTAYIDIPETELEIDSF 900 | |
| Qy | 911 VDACEWKVSDIRSGHFRLLRGWLQSLVQDTKLHLQNIHLHEPNRGKLAQYFAMNKDKKRR 970 | |
| Db | 901 VDACEWKVSDIRSGHFRLLRGWLQSLVQDTKLHLQNIHLHEPNRGKLAQYFAMNKDKKRR 960 | |
| Qy | 971 FKRESLPEORSQMGAFDRDAEYIRALRRHLVYYPARSAPVLSLSSGSWTLKLHFFQEV 1030 | |
| Db | 961 FKRESLPEORSQMGAFDRDAEYIRALRRHLVYYPARSAPVLSLSSGSWTLKLHFFQEV 1020 | |
| Qy | 1031 GASMRALYIVRDPRAWIYMLYNSKPSLYSLKNVPEHLAKLFKIEGGKGNLNSGYAPE 1090 | |
| Db | 1021 GASMRALYIVRDPRAWIYMLYNSKPSLYSLKNVPEHLAKLFKIEGGKGNLNSGYAPE 1080 | |
| Qy | 1091 YEPLRKELSKSKSNVSLLSHLWLANATAALRINTDLLPTSQYLVKFEDI VHFPOKTTTER 1150 | |
| Db | 1081 YEPLRKELSKSKSNVSLLSHLWLANATAALRINTDLLPTSQYLVKFEDI VHFPOKTTTER 1140 | |
| Qy | 1151 IPAFYIGLIPSLASNLQIIFATSTNLFYLPYEGEISPTNTNVWKQLPRDEIKLIENICWT 1210 | |
| Db | 1141 IPAFYIGLIPSLASNLQIIFATSTNLFYLPYEGEISPTNTNVWKQLPRDEIKLIENICWT 1200 | |
| Qy | 1211 LMDRLGYPKFMD 1222 | |
| Db | 1201 LMDRLGYPKFMD 1212 | |
| RESULT 4 | | |
| AA72643 | | |
| ID | AA72643 standard; protein; 1207 AA. | |
| XX | | |
| AC | AA72643; | |
| XX | | |
| DT | 02-MAY-2001 (first entry) | |
| XX | | |
| DE | Mouse glycosyl sulfotransferase-6 (GST-6). | |
| XX | | |
| KW | Mouse; glycosyl sulfotransferase-6; GST-6; immunosuppressive; therapy; | |
| KW | selectin binding inhibitor; gene therapy; inflammation; | |
| KW | systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes; | |

polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis; glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adenitis; Hashimoto's disease; Grave's disease; hypoparathyroidism; anaeimia; demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis; myocarditis; adult respiratory distress syndrome; eczema; psoriasis; asthma; hypersensitivity; rheumatic fever; tissue rejection.

Mus sp.
W0200106015-A1.
25-JAN-2001.
19-JUL-2000; 2000MO-US019741.
20-JUL-1999; 99US-0144694P.
13-JUN-2000; 2000US-00593828.
(REGC) UNIV CALIFORNIA.
Rosen SD, Lee JK, Hemmerich S;
WPI; 2001-138471/14.
N-PSDB; AAD02705, AAD02706.
New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for diagnostic and therapeutic agent screening applications.
Example 2; Fig 6B; 128pp; English.
The present sequence is mouse glycosyl sulfotransferase-6 (GST-6). GST is a type 2 membrane protein useful for inhibiting a binding event between a selectin and a selectin ligand, which comprises contacting the selectin with a non-sulphated selectin ligand, GST and a small molecular agent that inhibits the sulphation activity of GST. GST is also useful in inhibiting a selectin mediated binding event. GST is useful in gene therapy to treat disorders such as acute or chronic inflammation, systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes, glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's disease, Grave's disease, adenitis, hypoparathyroidism, pernicious anaemia, demyelinating diseases, cirrhosis, ulcerative colitis, dermatitis, myocarditis, regional enteritis, adult respiratory distress syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis, bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection during transplantation

Query Match 90.3%; Score 5899.5; DB 4; Length 1207;
Best Local Similarity 89.4%; Pred. No. 0;
Matches 1084; Conservative 68; Mismatches 55; Indels 5; Gaps 1;

QY 11 MALMFTGHLALFALLMPAFSTFEESVNSYSEWAVFTDDIDQFKTKQVQDFRPQKLKSM 70
DB 1 MAPNTEHLFLFLMCCSFSTCEBSVNSYSEWAVFTDDIQWLKSQIQDFKLR----- 55
QY 71 LHPSLYFDAGEIQAMRQKSRASHLHLFRATRSVTVMLSNPTYYLPPLPPKADFAAKWNEI 130
DB 56 LHPNLVFDAGDIQTLKQKSRTHLHIFRAIKSAVTIMLSNPSYYLPPLPPKAEFAAKWNEI 115
QY 131 YGNLPLALYCLCPEDKVAFEVLEVMQVYKDWLVENAPGDEVPIGHSITGFATA 190
DB 116 YGNLPLALYCLCPEDKVAFEVLEVMQVYKDWLVENAPGDEVPIGHSITGFATA 175
QY 191 FDFLYNLNDHRRQKYLEKIWIITEEMVEYKVRSGWQKQLLHNHQATNMIALLTGALVTG 250
DB 176 FDFLYNLGNQRQKYLEKIWIITEEMVEYKVRSGWQKQLLHNHQATNMIALLTGALVTG 235
QY 251 VDKGSKANIWKQAVVDVNEKTMFLNHNHIVDGSLSDEGVAYSYTAQSVQYVFLAQRHFI 310
DB 236 VDKGSKANIWKQAVVDVNEKTMFLNHNHIVDGSLSDEGVAYSYTAQSVQYVFLAQRHFI 295

QY 311 NNLDNNWLKQHFYFATYATLLPGFORTVGIADSNYNWYFPGESQVLFDKFLKNGAGNWL 370
DB 296 NNFDNNWLKQHFYFATYATLLPGFORTVGIADSNYNWYFPGESQVLFDKFLKNGAGNWL 355
QY 371 AQOIRKRPKQGMVPSSTAQRSTLHTEYIWDVDPQLTPOPPADYGTAKIHTFPNMGVVY 430
DB 356 AQOIRKRPKQGMVPSSTAQRSTLHTEYIWDVDPQLTPOPPADYGTAKIHTFPNMGVVY 415
QY 431 GAGLPNTQNTTFVSFKSGKLGGRVAVYDI VHFQPSYWDGWRSPNPGHEHPDQNSFTFAPN 490
DB 416 GGGLPNTQNTTFVSFKSGKLGGRVAVYDI VHFQPSYWDGWRSPNPGHEHPDQNSFTFAPN 475
QY 491 GQVFSEALYGPKLSHLNNVLVFPSPSSQCNKPWEGQLGCEAQWLKWTGEEVGDAGEI 550
DB 476 GQVFSEALYGPKLSHLNNVLVFPSPSSQCNKPWEGQLGCEAQWLKWTGEEVGDAGEV 535
QY 551 ITASQHGEMVYSGEAVSAYSSAMRLKSVYRALLLNSQTLVVDHIERODESPINSVA 610
DB 536 ITAAQHGDRMFVSGEAVSAYSSAMRLKSVYRALLLNSQTLVVDHIERODESPINSVA 595
QY 611 FPHNLIDFKYIPYKPMNRYNGAMDMVDWAHYKMFWDHGNSPMASIQEAEQAAEFKKR 670
DB 596 FPHNLIDFKYIPYKPMNRYNGAMDMVDWAHYKMFWDHGNSPMASIQEAEQAAEFKKR 655
QY 671 WTOPVNVTFOMESTITRIAYVYFVGYPIVNVSSCRPIDSSNPGLOISLNNVNTHEHVSIVTD 730
DB 656 WTOPVNVTFOMESTITRIAYVYFVGYPIVNVSSCRPIDSSNPGLOISLNNVNTHEHVSIVTD 715
QY 731 YHNLKTRFNYLGGFGFASVADOGITRFGIGTOAIYKVPVRHRIIPFGPKENIAGLIL 790
DB 716 YQNLKSRFVYLGFGFASVANOGQITRFGIGTOAIYKVPVRHRIIPFGPKENIAGLIL 775
QY 791 CISLVILTQWRFLYLFKRLMRMLILVIALWPIELLDVNMSTCSQPTCAKWTREABSK 850
DB 776 CISLVILTQWRFLYLFKRLMRMLILVIALWPIELLDVNMSTCSQPTCAKWTREABSK 835
QY 851 KSLSSGHHMDLPDVVITSLPGSGAILKQLPFNSDFLYRVPTAYIDIPETELEDSP 910
DB 836 KWLSEGHVLDLPNVITSLPGSGAILKQLPFNSDFLYRVPTAYIDIPETELEDSP 895
QY 911 VDAKWKVSDIRSGHFLRLGWLQSLVODTKLHLQNLHLEPNRGLAQYFAMNKDKKR 970
DB 896 VDAKWKVSDIRSGHFLRLGWLQSLVODTKLHLQNLHLEPNRGLAQYFAMNKDKKR 955
QY 971 FKRESLPEORSOMKGAFFORDAEYIRALRHLVYPSARVLSLSSGSWTLKLFHFOEVL 1030
DB 956 LKRESLQDQORSIKGPFDRDAEYIRALRHLVYPSARVLSLSSGSWTLKLFHFOEVL 1015
QY 1031 GASMRALYIVRPRAMIYSLNYSKPSLYSLKNVPEHLAKLFKIEGKGKCNLSGYAFE 1090
DB 1016 GTSMRALYIVRPRAMIYSLNYSKPSLYSLKNVPEHLAKLFKIEGKGKCNLSGYAFE 1075
QY 1091 YEPRLKELSKSNVSLSHLWLANATAALRINTDLPSTSYOLVKFEDIVHFPKQKTER 1150
DB 1076 YESLKELEISQSNALISLHSLWLANATAALRINTDLPSTSYOLVKFEDIVHFPKQKTER 1135
QY 1151 IFAPLGIPLSPASINOLIFATSNLFLPYEGEISPTNTNWKONLPRDEIKLIENICWT 1210
DB 1136 IFAPLGIPLSPASINOLIFATSNLFLPYEGEISPTNTNWKONLPRDEIKLIENICWT 1195
QY 1211 LMDRLGYPKFMD 1222
DB 1196 LMDHLGYPKFMD 1207
RESULT 5
AAE33542
ID AAE33542 standard; protein; 1207 AA.
XX AAE33542;
AC AAE33542;
XX
DT 16-APR-2003 (first entry)
XX

DE Human novel CpG-associated gene 1 (NCAG1) encoded protein #2.
 KW Human; novel CpG-associated gene 1; bipolar disorder; neuroprotective;
 KW NCAG1; mood disorder; chromosome 18.
 XX
 OS Homo sapiens.
 XX WO2002101044-A2.
 XX
 XX 19-DEC-2002.
 XX
 XX 06-JUN-2002; 2002WO-EF006316.
 XX
 XX 11-JUN-2001; 2001EP-00202214.
 XX (JANC) JANSSEN PHARM NV.
 XX Del-Favero JPL, Van Broeckhoven C;
 XX WPI; 2003-148807/14.
 XX N-PSDB; AAD50033.
 XX
 PT New brain expressed genes (designated novel CpG-associated Gene 1
 PT (NCAG1)) and its encoded protein, useful as diagnostic markers for
 PT bipolar or mood disorders, and as targets for developing drugs for the
 PT treatment these disorders.
 XX
 XX Claim 16; Col 48-51; 26pp; English.
 XX
 CC The invention relates to novel CpG-associated gene 1 (NCAG1) brain-
 CC expressed gene and its encoded protein. The NCAG1 nucleic acid or its
 CC encoded protein is useful as a diagnostic marker for bipolar disorder
 CC such as mood disorders. They are also useful as targets for developing
 CC drugs, as well as for target validation, for the treatment of bipolar
 CC disorders. The present sequence is human NCAG1 protein. NCAG1 gene is
 CC located at chromosome 18
 XX
 SQ Sequence 1207 AA;
 Query Match 90.3%; Score 5899.5; DB 6; Length 1207;
 Best Local Similarity 89.4%; Pred. No. 0;
 Matches 1084; Conservative 68; Mismatches 55; Indels 5; Gaps 1;
 QY 11 MALMFTGHLFLALLMFAFTFESVNTSEWAVFTDDIDQFTQKQVDFRPNQKLKSM 70
 DB 1 MAFMFTHEHLFLTLMMCSFTCESVNTSEWAVFTDDIQWLKSKQIQDFKLNRR----- 55
 QY 71 LHPSLYFDAGEIQAMRQKGRASHLHLFRAIRSAVTVMLSNPTYLPPPPKADFAAKWNEI 130
 DB 56 LHPNLVFDAGDIQTLKQKSRSTSHLHIFRAIKSAVTIMLSNPSTYLLPPPKHAEFAAKWNEI 115
 QY 131 YGNLPLALYCLLPEDKVAFFEVLYEDRMVGYKDWLVENAPGDVEPIGHSLTGFATA 190
 DB 116 YGNLPLALYCLLPEDKVAFFEVLYEDRMVGYKDWLVENAPGDVEPIGHSLTGFATA 175
 QY 191 FDFLYLLDNHRRQKYLEKIWLITEWMEYKSVRSWKGKOLLHNHQAATNMIALLTGALVTG 250
 DB 176 FDFLYLLNGLNQRQKYLEKIWLITEWMEYKSVRSWKGKOLLHNHQAATNMIALLIGALVTG 235
 QY 251 VDKGSKANIWKQAVDVDMKTMFLNHI VDGSLDEGVAYGYSYAKSVTYQVFLAQRHFN 310
 DB 236 VDKGSKANIWKQAVDVDMKTMFLNHI VDGSLDEGVAYGYSYAKSVTYQVFLAQRHFN 295
 QY 311 NNLDNNLWKHFWFYFATLLPGFQRTVGTADSNYNWYFPGESQVLFDKFLKNGAGNWL 370
 DB 296 NNFDNNLWKHFWFYFATLLPGYQRTVGTADSNYNWYFPGESQVLFDKFLKNGAGNWL 355
 QY 371 AQQIRKRPKDGPMWSTAQRWSTLHTEYIWDYDQTLTPOPPADYGTAKHTFPNMGVVTY 430
 DB 356 AQQIRKRPKDGPMWSTAQRWSTLHTEYIWDYDQTLTPOPPADYGTAKHTFPNMGVVTY 415
 QY 431 GAGLPNTQNTFVSFGSKLGGRAVYDIVHFQYPSWIDGWRSPNPGHEHPDQNSFTFAPN 490

DB 416 GGGLPNTQNTFVSFGSKLGGRAVYDIVHFQYPSWIDGWRSPNPGHEHPDQNSFTFAPN 475
 QY 491 GQVFVSEALYGPKLSHLANNVLVFAFSPSSQCNKPMWEGOLGECAGQWLKWTGBEVBGAAGRI 550
 DB 476 GQVFVSEALYGPKLSHLANNVLVFAFSPSSQCNKPMWEGOLGECAGQWLKWTGBEVBGAAGRI 535
 QY 551 ITASQHGEMVFSGEAVSAYSSAMRLKSVYRALLLNSQTLVVDVHIERQEDSPINSVSA 610
 DB 536 ITAAQHGDRMFVSGEAVSAYSSAMRLKSVYRALLLNSQTLVVDVHIERQEDSPINSVSA 595
 QY 611 PFHNLIDIDFKYIPYKFMRYNGAMMDVMDAHYKMFWDHGHGNSPMASIOEABQAAEFKKR 670
 DB 596 PFHNLIDIDFKYIPYKFMRYNGAMMDVMDAHYKMFWDHGHGNSPMASIOEABQAAEFKKR 655
 QY 671 WQFVNVTQFMESTITRIAYVYFGYINVSRCRFDSSNPGLOISLVNNTNTEHVSIVTD 730
 DB 656 WQFVNVTQFMESTITRIAYVYFGYINVSRCRFDSSNPGLOISLVNNTNTEHVSIVTD 715
 QY 731 YHNLKTRFNLYGFGFASVADQGITRFGGLGTQAIKVPVRHDIRIIFPGFKFNIAVGIL 790
 DB 716 YQNLKSRFSYLGFGFASVADQGITRFGGLGTQAIKVPVRHDIRIIFPGFKFNIAVGIL 775
 QY 791 CISLVILTFQWRFYLSFRKLWRWILVIALWFIELLDVWSTCSQPICAKWTTRTBAESK 850
 DB 776 CISLVILTFQWRFYLSFRKLWRWILVIALWFIELLDVWSTCSQPICAKWTTRTBAESK 835
 QY 851 KSLSEGHMMDLPDVVITSLPGSGAEILKQLFNSSDFLYIRVPTAYIDIPETELEIDSF 910
 DB 836 KVMISEGHVVDLPNVIIITSLPGSGAEILKQLFNSSDFLYIRVPTAYIDIPETELEIDSF 895
 QY 911 VDACEWKVSDIRSGHFRLLRGLQSLVQDTKLHLQNIHLHFNPKGLAQYFAMNKDKRK 970
 DB 896 VDACEWKVSDIRSGHFRLLRGLQSLVQDTKLHLQNIHLHFNPKGLAQYFAMNKDKRK 955
 QY 971 FKRRSLPEQSQMKGAFDRDAEYIRALRHLVYVPSARPVLSLSSGWTILKHFQEV 1030
 DB 956 LKRRESLQDRSRINKGPPDRDAEYIRALRHLVYVPSARPVLSLSSGWTILKHFQEV 1015
 QY 1031 GASMRALYIVRDPRAWIYSLMYNKPSTLYSLKNVPEHLAKLFKEGGKCNLNSGYAFE 1090
 DB 1016 GTSMRALYIVRDPRAWIYSLMYNKPSTLYSLKNVPEHLAKLFKEGGKCNLNSGYAFE 1075
 QY 1091 YEPLRKELSKSNASVLSLHLWLAATAALRINTDLPTSYQLVKPEDI VHFPOKTTTER 1150
 DB 1076 YESLKELEISQSNALSLSHLWLAATAALRINTDLPTSYQLVKPEDI VHFPOKTTTER 1135
 QY 1151 IPAFILGIPLSPASLNQILFATSTNL FYLPYGEISPTNWNWKNLPRDEIKLIENICWT 1210
 DB 1136 IPAFILGIPLSPASLNQILFATSTNL FYLPYGEISPTNWNWKNLPRDEIKLIENICWT 1195
 QY 1211 LMDRLGYPKFMD 1222
 DB 1196 LMDHLGYPKFMD 1207
 RESULT 6
 AAB93735
 ID AAB93735 standard; protein; 755 AA.
 XX
 AC AAB93735;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human protein sequence SEQ ID NO:13382.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 XX Homo sapiens.
 XX EP1074617-A2.
 XX
 PD 07-FEB-2001.
 XX

28-JUL-2000; 2000EP-00116126.
 29-JUL-1999; 99JP-00248036.
 27-AUG-1999; 99JP-00300253.
 11-JAN-2000; 2000JP-00118776.
 02-MAY-2000; 2000JP-00183767.
 09-JUN-2000; 2000JP-00241899.
 (HELI-) HELIX RES INST.
 Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 Iehii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 WPI; 2001-318749/34.
 Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.
 Claim 8; SEQ ID NO 13382; 2537pp + Sequence Listing; English.
 The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises:
 (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention
 Query Match
 Best Local Similarity 61.8%; Score 4037; DB 4; Length 755;
 Matches 750; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 217 MYEYKRVSWGKQLLNHQAATNIALTGALVTGVDGSKANIKWQAVVDVMEKTMFLN 276
 1 MYEYKRVSWGKQLLNHQAATNIALTGALVTGVDGSKANIKWQAVVDVMEKTMFLN 60
 277 HIVDGLDEGVAYGYSYAKSTQVVELAQRHFNINLNNLWKLKHFYFYATLPGFQRT 336
 61 HIVDGLDEGVAYGYSYAKSTQVVELAQRHFNINLNNLWKLKHFYFYATLPGFQRT 120
 337 VGIADSNYNWFYGPESQVFLDKFKLKNAGNWLAAQIRKHPKDGPMVPTAQRWSTLH 396
 121 VGIADSNYNWFYGPESQVFLDKFKLKNAGNWLAAQIRKHPKDGPMVPTAQRWSTLH 180
 397 TEYIWDYDQLPQPPADYGTAKIHFTFPNMGVVTYAGLPTNTQNTVFSFKSGKLGGRVY 456
 181 TEYIWDYDQLPQPPADYGTAKIHFTFPNMGVVTYAGLPTNTQNTVFSFKSGKLGGRVY 240
 457 DIVHFQPSWIDGWRSPNPGHEHPDQNSFTAPNGQVFSVSAALYGPKLHNNLVFAPS 516
 241 DIVHFQPSWIDGWRSPNPGHEHPDQNSFTAPNGQVFSVSAALYGPKLHNNLVFAPS 300
 517 PSSQCNKPEWGLGCAQWLKWTGEEVGDAAAGIITASQHGEMVFSVSGEAVSAYSSAMRL 576

301 PSQCNKPEWGLGCAQWLKWTGEEVGDAAAGIITASQHGEMVFSVSGEAVSAYSSAMRL 360
 577 KSVYRALLLNSQTLVVDHIERQEDSPINSVSAFFHNLDIDFKYIPYKPMNRYNGAMMD 636
 361 KSVYRALLLNSQTLVVDHIERQEDSPINSVSAFFHNLDIDFKYIPYKPMNRYNGAMMD 420
 637 VMDAHYKMFWDHGHNSPMASIOEAQAAEFKRGWTQFVNVTFQMESTIIRIAYVFYGPY 696
 421 VMDAHYKMFWDHGHNSPMASIOEAQAAEFKRGWTQFVNVTFQMESTIIRIAYVFYGPY 480
 697 INVSSCRFDSSNPGLOISLVNNTNTHVSVIVTDYHNLKTRNVLGFGGSPASVADQOQIT 756
 481 INVSSCRFDSSNPGLOISLVNNTNTHVSVIVTDYHNLKTRNVLGFGGSPASVADQOQIT 540
 757 RFLGIGTQAIKVRHRIIPFPGFKFNIAVGLILCISLVILTQWRFYLSFRKLMRWILI 816
 541 RFLGIGTQAIKVRHRIIPFPGFKFNIAVGLILCISLVILTQWRFYLSFRKLMRWILI 600
 817 LVIALWFIELLDVWSTCSQPICAKWTRTAEAGSKLSSEGHHMDLPDVVITSLPGSGAE 876
 601 LVIALWFIELLDVWSTCSQPICAKWTRTAEAGSKLSSEGHHMDLPDVVITSLPGSGAE 660
 877 ILKQLFNSDFLYIRVPTAYIIDIPETELIDSFVDACEKVDIRSGHFLRGLWLOSL 936
 661 ILKQPFNSDFLYIRVPTAYIIDIPETELIDSFVDACEKVDIRSGHFLRGLWLOSL 720
 937 VQDTKLHLQNIHLHPEHNRGLAOYFAMNKDKKKK 970
 721 VQDTKLHLQNIHLHPEHNRGLAOYFAMNKDKKKK 754
 RESULT 7
 ADP56673
 ID ADP56673 standard; protein; 755 AA.
 XX AC ADP56673;
 XX DT 12-AUG-2004 (first entry)
 XX DE Human protein which is a chondroitin 6 sulphotransferase 3 homologue.
 XX KW cancer detection; large intestine; oesophagus; stomach; lungs; pancreas;
 XX KW liver; kidney; colon; human; chondroitin 6 sulphotransferase 3; CH6T3.
 XX OS Homo sapiens.
 XX PN JP2004147505-A.
 XX PD 27-MAY-2004.
 XX PF 28-OCT-2002; 2002JP-00312927.
 XX PR 28-OCT-2002; 2002JP-00312927.
 XX PA (DOKU-) DOKURITSU GYOSEI HOJIN SANGYO GIJUTSU SO.
 XX PA (FURE) FUJIREBIO KK.
 XX PA (SEGG) SEIKAGAKU KOGYO CO LTD.
 XX WI WPI; 2004-434537/41.
 XX DR N-PSDB; ADP56672.
 XX PT Novel nucleic acid such as DNA, useful as probe for identifying or
 PT detecting cancerous tissue of esophagus, stomach, lungs, pancreas, liver,
 PT kidney or colon, preferably large intestine tissue.
 XX PS Example 1; SEQ ID NO 2; 24pp; Japanese.
 XX CC The invention relates to a novel nucleic acid having 40-1000 base pairs
 CC and comprising a sequence which is complementary to 41-2308 nucleotides
 CC of a fully defined sequence of 3288 base pairs as given in the
 CC specification. The polynucleotide of the invention may be useful for
 CC rapid and highly reliable detection of cancerous tissue derived from the
 CC large intestine, oesophagus, stomach, lungs, pancreas, liver, kidney or

CC colon. The current sequence is that of the human protein of the invention
CC which is a chondroitin 6 sulphotransferase 3 (Ch6T3) homologue.

XX
SQ Sequence 755 AA;

Query Match 61.8%; Score 4037; DB 8; Length 755;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 750; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 217 MYEYSKVRSGKQLLNHQTNNIALLTGALVTGVDKSGKANIWKQAVDVMKTMFLN 276
DB 1 MYEYSKVRSGKQLLNHQTNNIALLTGALVTGVDKSGKANIWKQAVDVMKTMFLN 60
QY 277 HIVDGSLEDEGVAGSYTAKSVTVQVFLAQRHENINLNNLWKLWHEWYATLLPGFQRT 336
DB 61 HIVDGSLEDEGVAGSYTAKSVTVQVFLAQRHENINLNNLWKLWHEWYATLLPGFQRT 120
QY 337 VGIADSNYNFYGPESQLVFLDKFILKNGAGNLAQQIRKHPKDGMPVPSTAQRMSTLH 396
DB 121 VGIADSNYNFYGPESQLVFLDKFILKNGAGNLAQQIRKHPKDGMPVPSTAQRMSTLH 180
QY 397 TEIWIYDQPLTPPPADYGTAKIHTFPNWGVVYTGAGLPNTQNTVSPKSGKLGRAVY 456
DB 181 TEIWIYDQPLTPPPADYGTAKIHTFPNWGVVYTGAGLPNTQNTVSPKSGKLGRAVY 240
QY 457 DIVHFQPSYWDGWRSPNPGHEHPDQNSFTFAPNGQVFVSEALYGPKLHLNNVLFPAS 516
DB 241 DIVHFQPSYWDGWRSPNPGHEHPDQNSFTFAPNGQVFVSEALYGPKLHLNNVLFPAS 300
QY 517 PSSQCNKPEWGQCECAQWLKWTGEEVGDAAGIITASQHGEMVFSGEAVSAYSSAMRL 576
DB 301 PSSQCNKPEWGQCECAQWLKWTGEEVGDAAGIITASQHGEMVFSGEAVSAYSSAMRL 360
QY 577 KSVYRALLLNSQTLVVDHIERQEDSPINSVAFPHNLIDIDPKYIPYKPMNRNGAMMD 636
DB 361 KSVYRALLLNSQTLVVDHIERQEDSPINSVAFPHNLIDIDPKYIPYKPMNRNGAMMD 420
QY 637 VMDAHYKMFDFDHGNSPMASIOEAQAEFKKRWTFQVNVTFQMESTITRIAYVFGPY 696
DB 421 VMDAHYKMFDFDHGNSPMASIOEAQAEFKKRWTFQVNVTFQMESTITRIAYVFGPY 480
QY 697 INVSSCRFIDSSNPGQLQISLVNNTNTHVSVIVTDYHNLKTRFNLYLGGFGFASVADQGIT 756
DB 481 INVSSCRFIDSSNPGQLQISLVNNTNTHVSVIVTDYHNLKTRFNLYLGGFGFASVADQGIT 540
QY 757 RFLGTOAIKVPVRHRIIPFGKFNIAVGLILCISLVILTPQWRFPYLSFRKLMRWILI 816
DB 541 RFLGTOAIKVPVRHRIIPFGKFNIAVGLILCISLVILTPQWRFPYLSFRKLMRWILI 600
QY 817 LVIALWFIELLDVWSTCSQPICAKWRTEAGSKSLSSGHHMDLPDVVITSLPGSGAE 876
DB 601 LVIALWFIELLDVWSTCSQPICAKWRTEAGSKSLSSGHHMDLPDVVITSLPGSGAE 660
QY 877 ILKQLFNNSDFLYIRVPTAYIDIPETELEIDSFVDACEWKVSDIRSGHFRLLRGLQSL 936
DB 661 ILKQLFNNSDFLYIRVPTAYIDIPETELEIDSFVDACEWKVSDIRSGHFRLLRGLQSL 720
QY 937 VQDTKLHLQNLHLEPNRGKLAQYFAMNKDKKK 970
DB 721 VQDTKLHLQNLHLEPNRGKLAQYFAMNKDKKK 754

RESULT 8
AA72641
ID AAY72641 standard; protein; 596 AA.
XX
AC AAY72641;
XT 02-MAY-2001 (first entry)
XX Human glycosyl sulfotransferase-6 (GST-6) fragment.
XX Human; glycosyl sulfotransferase-6; GST-6; immunosuppressive; therapy;

KW selectin binding inhibitor; gene therapy; inflammation;
KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
KW Polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
KW Glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenalitis;
KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
KW asthma; hypersensitivity; rheumatic fever; tissue rejection.

XX Homo sapiens.
OS WO200106015-A1.

XX 25-JAN-2001.
PD 19-JUL-2000; 2000WO-US019741.

XX 20-JUL-1999; 99US-0144694P.
PR 13-JUN-2000; 2000US-00593828.

XX (REGC) UNIV CALIFORNIA.
PA Rosen SD, Lee JK, Hemmerich S;
PI WPI; 2001-138471/14.

XX N-PSDB; AAD02702, AAD02703.
DR New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
XX diagnostic and therapeutic agent screening applications.

XX Claim 3; Fig 3; 128pp; English.
PS The present sequence is human glycosyl sulfotransferase-6 (GST-6)

CC fragment. GST is a type 2 membrane protein useful for inhibiting a
CC binding event between a selectin and a selectin ligand, which comprises
CC contacting the selectin with a non-sulphated selectin ligand, GST and a
CC small molecular agent that inhibits the sulphation activity of GST. GST
CC is also useful in inhibiting a selectin mediated binding event. GST is
CC useful in gene therapy to treat disorders such as acute or chronic
CC inflammation, systemic lupus erythematosus (SLE), rheumatoid arthritis,
CC polyarteritis nodosa, polymyositis, dermatomyositis, systemic sclerosis,
CC diabetes, glomerulonephritis, myasthenia gravis, Sjogren's syndrome,
CC Hashimoto's disease, Grave's disease, adrenalitis, hypoparathyroidism,
CC pernicious anaemia, demyelinating diseases, cirrhosis, ulcerative
CC colitis, dermatitis, myocarditis, regional enteritis, adult respiratory
CC distress syndrome, infantile eczema, psoriasis lichen planus, allergic
CC rhinitis, bronchial asthma, hypersensitivity, rheumatic fever and tissue
CC rejection during transplantation

SQ Sequence 596 AA;
Query Match 48.2%; Score 3147; DB 4; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.3e-278;
Matches 596; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 627 MNRNGAMMDVMDAHYKMFDFDHGNSPMASIOEAQAEFKKRWTFQVNVTFQMESTIT 686
DB 1 MNRNGAMMDVMDAHYKMFDFDHGNSPMASIOEAQAEFKKRWTFQVNVTFQMESTIT 60

QY 687 RIAYVFGPYINVSRCRFDSSNPGQLQISLVNNTNTHVSVIVTDYHNLKTRFNLYLGGFG 746
DB 61 RIAYVFGPYINVSRCRFDSSNPGQLQISLVNNTNTHVSVIVTDYHNLKTRFNLYLGGFG 120

QY 747 ASVADQGITRFGLTQAIKVPVRHRIIPFGKFNIAVGLILCISLVILTPQWRFPYLS 806
DB 121 ASVADQGITRFGLTQAIKVPVRHRIIPFGKFNIAVGLILCISLVILTPQWRFPYLS 180

QY 807 FRKLMRWILILVIALWFIELLDVWSTCSQPICAKWRTEAGSKSLSSGHHMDLPDVV 866
DB 181 FRKLMRWILILVIALWFIELLDVWSTCSQPICAKWRTEAGSKSLSSGHHMDLPDVV 240

QY 867 ITSLPGSGAEILKQLFNNSDFLYIRVPTAYIDIPETELEIDSFVDACEWKVSDIRSGH 926

241 ITSPLGSGABILKQLFFNSDFLYIRVPTAYIDIPETELEIDSFVDACEWVKVDIRSGHF 300
927 RLARGWLQSLVQDTKLHQLHLEHNRGKLAQYFANWKDKRKFGRRESLPEQSRQMG 986
301 RLARGWLQSLVQDTKLHQLHLEHNRGKLAQYFANWKDKRKFGRRESLPEQSRQMG 360
987 AFDRDAEYIRALRRHLVYPSARPVLSSSGSWTLKLHFFQEVLGASMRALYIVRDPRAW 1046
361 AFDRDAEYIRALRRHLVYPSARPVLSSSGSWTLKLHFFQEVLGASMRALYIVRDPRAW 420
1047 IYMLYNSKPSLSLVKQVPEHLAKLFKEGKGKCNLSNGYAFYEYEPRLKELSKSNAV 1106
421 IYMLYNSKPSLSLVKQVPEHLAKLFKEGKGKCNLSNGYAFYEYEPRLKELSKSNAV 480
1107 SLLSHLWLANTAALRINTDLPSTYQLVKFPEDIVHPQKTERIFAPLGIPLSPASLNQ 1166
481 SLLSHLWLANTAALRINTDLPSTYQLVKFPEDIVHPQKTERIFAPLGIPLSPASLNQ 540
1167 ILFATSTNLFYLPYEGEISPTNTNVWKQNLPRDEIKLIENICWTLMDRLGYPKFMD 1222
541 ILFATSTNLFYLPYEGEISPTNTNVWKQNLPRDEIKLIENICWTLMDRLGYPKFMD 596

RESULT 9
AAB94410
ID AAB94410 standard; protein; 480 AA.
XX
AC AAB94410;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:14997.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EPI074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-00116126.
XX
PR 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.
XX
PS Claim 8; SEQ ID NO 14997; 2537pp + Sequence Listing; English.
XX
CC The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises:
CC (a) an oligo-dr primer and an oligonucleotide complementary to the
CC complementary strand of a polynucleotide which comprises one of the 5602
CC nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the

oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAB03166 to AAB13628 and AAB13633 to AAB18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAB13629 to AAB13632 represent oligonucleotides, all of which are used in the exemplification of the present invention
XX
SQ Sequence 480 AA;
Query Match 39.8%; Score 2597; DB 4; Length 480;
Best Local Similarity 99.8%; Pred. No. 2.4e-228;
Matches 479; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 320 MHFMYATLLPGFQRTVGIADSNYNWFGPESQVFLDKFILKNGAGNWLAAQIRKHP 379
DB 1 MHFMYATLLPGFQRTVGIADSNYNWFGPESQVFLDKFILKNGAGNWLAAQIRKHP 60
QY 380 KDGMPVPTAQRWSTLHTEYIWDPOLTPQPPADYGTAKIHTFPNMGVVTYAGLPNTQT 439
DB 61 KDGMPVPTAQRWSTLHTEYIWDPOLTPQPPADYGTAKIHTFPNMGVVTYAGLPNTQT 120
QY 440 NTFVSPKSGKLGGRVYDIVHFPYSWIDGWRSPNPGHEHPDQNSPTTFAPNGOVFVSEAL 499
DB 121 NTFVSPKSGKLGGRVYDIVHFPYSWIDGWRSPNPGHEHPDQNSPTTFAPNGOVFVSEAL 180
QY 500 YGPKLSHLNNLVFAPSPSQCNKPWEGQLGECQAQMLKWTGEEVGDAAAGIITASQHGEM 559
DB 181 YGPKLSHLNNLVFAPSPSQCNKPWEGQLGECQAQMLKWTGEEVGDAAAGIITASQHGEM 240
QY 560 VFVSGEAVSAYSSAMRLKSVYRALLNLSQTLVVDHIERQEDSPINSVSAFFHNLDIDF 619
DB 241 VFVSGEAVSAYSSAMRLKSVYRALLNLSQTLVVDHIERQEDSPINSVSAFFHNLDIDF 300
QY 620 KYTPYKPMRYNGAMDMVDADAHYKMFWDHGHNSPMASIOEASQAAEFKKRWTFQVNVTF 679
DB 301 KYTPYKPMRYNGAMDMVDADAHYKMFWDHGHNSPMASIOEASQAAEFKKRWTFQVNVTF 360
QY 680 QMESTITRIAYFYGPYINVSRCRFDSSNPGIQLISLVNNTTEHVSVITVDYHNLKTRFN 739
DB 361 QMESTITRIAYFYGPYINVSRCRFDSSNPGIQLISLVNNTTEHVSVITVDYHNLKTRFN 420
QY 740 YLGGFGFASVADQGITRFGLTQAIKVPVRHRIIPPGFKFNIAVGLILCISLVILTF 799
DB 421 YLGGFGFASVADQGITRFGLTQAIKVPVRHRIIPPGFKFNIAVGLILCISLVILTF 480

RESULT 10
AAY51120
ID AAY51120 standard; protein; 958 AA.
XX
AC AAY51120;
XX
DT 24-MAR-2000 (first entry)
XX
DE Human SART-2 protein.
XX
KW SART-2; human; tumor antigen protein; T cell; MHC class I antigen;
KW major histocompatibility complex.
XX
OS Homo sapiens.
XX
PN JPI1318455-A.
XX
PD 24-NOV-1999.
XX
PF 08-MAY-1998; 98JP-00126398.
XX

```

PR 08-MAY-1998; 98JP-00126398.
XX (ITOY/) ITO Y.
XX WPI; 2000-090523/08.
DR N-ESDB; AAZ44181.
XX Human cancer involution antigen protein.
XX Claim 1; Page 11-15; 19pp; Japanese.
XX This invention describes a novel human tumor antigen (SART-2)
CC recognized by T cells by combining with major histocompatibility complex
CC (MHC) class I antigens. The tumor antigen protein can be used for the
CC treatment and the diagnosis of tumor cells other than melanoma cell,
CC particularly canceroid. This sequence represents the human SART-2 protein
CC described in the method of the invention,
XX
XX Sequence 958 AA;
XX
XX Query Match 26.8%; Score 1753.5; DB 3; Length 958;
XX Best Local Similarity 49.0%; Pred. No. 1.5e-150;
XX Matches 341; Conservative 114; Mismatches 210; Indels 31; Gaps 9;
XX
QY 72 HPSLYFDAGETQAMRQKSRASHLHLFRAIRSAVTVMLSNPTTYLPPPKHADFAAKWNEY 131
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 43 HPMYFSAEVAELQLRASASHEHIAARLTAHVHTMLSSPLEYLPWDPKDYSAKWEIF 102
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 132 GNNLPPLALYCLCPEDKVAPEFVLEVMYDMRVGKWLVENAFGDEVPIGHSLTGATFAT 191
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 103 GNNLGALAMFCVLYPENIEARDMAKDYMERMAAQPSMLVKDAPWDEVLASHLVGFATAY 162
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 192 DFLYNLLDNHRRQKYLEKIWIITEEMYEYSKVRSGKQLLNHNQATNMIALTGALVTGV 251
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 163 DFLYNLSKTKQKEFLEVIANASGYMYETSYRRGWFQYLLHNHQPNTCWMALLTGSVL-M 221
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 252 DKG--SKANTWKOAVDVDMKTFPLNHIVDGSLDEGVAYGVTAKSVTOYVFLAQHFN 309
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 222 NOGVLOEAYLWTKQVLTIMESLVLLREVTDGSLYEGVAGSYTTSLSLFOYMLVQRHFN 281
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 310 INNDNNLWKHFWFYATLLPGFQRTVGADSNYNWFYGPESQVLPDLKFLKNGAGNW 369
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 282 INHFGHPWLKQHFAMFYRTLLPGFQRTVAIDSNYNWFYGPESQVLPDLKFLKNGAGNW 341
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 370 LAQOIKRHPKDGMPVSTQARSTLHTEYIWDVDPQLTPQPPADYGTAKHTHTPNWGVVT 429
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 342 LADQIRNRNVVEGGTSPKGRWCCTLHTEFLWYDGLSKSVPPDPFGTPTLHYFEDRGVVT 401
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 430 YGAGLPNTQNTFVSFKSGKLGRAVDYIVHFQPY-SWIDGWSFNPGHEHPDQNSFTFA 488
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 402 YGSALPAEINRSFLSKGLGGRALYDVRNKYKDWIKGWNFNAGHEHPDQNSFTFA 461
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 489 PNGQVFSVEALYGPKLHNLVLPSPSSQCNKPWEGQLGE--CAQMLKWTGVEVGDA 546
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 462 PNGVPFTEALYGPKYTFNNVLMFSPAVSKSCFSPWVGQVTEDCSKSKYKHDLAASC 521
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 547 AGEIITASQHGEMVFSVGEAVSAYSSAMRLKSVYRALLILNSQTLVWDHIERQEDSPIN 606
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 522 QGRVVAEEKNGVVFIKRGEGVAGYNPOLNKNQVRLNLIHLPQLLLVQIHLGEESSPLE 581
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 607 SVSAFFHNLIDIFKYIPKFMNRYNGAMDMVDVAHYKMFDFHSGNSPMASIQEAQAAE 666
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 582 TAASFFHNVDPPE---ETVVDGVHGAFFIRQRLGLYKMYWMDTGTSEKATFASVYTPRG 638
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 667 FKRWTOFVNVTQMESTIRIAYVFGPYINVSSCRFDSSNPGIQLISLVNNTSHVVS 726
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 639 YPYNGTNYVNVTHLRSPITRAAYFLIGPSIDVQS-----FTVHGDSQQLDVF 686
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 727 IVDYHNLKTRFNVL-----GGGGFASV-ADOGQI 755
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 687 IATSKHAYAT---YLTWTEATGQSFAQVIAADRHKI 719
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :

```

```

RESULT 11
ADL83102
ID ADL83102 standard; protein; 958 AA.
XX
XX AC ADL83102;
XX
XX DT 17-JUN-2004 (first entry)
XX
XX DE Human PRO49564, SEQ ID 304.
XX
XX KW Immunosuppressive; Cytostatic; Antiarthritic; Antirheumatic; Antianemic;
XX Antiallergic; Muscular; Neuroprotective; Nephrotropic; Antiinflammatory;
XX Gene Therapy; PRO; B cell related disorder; cancer;
XX KW immune-mediated inflammatory disease; human.
XX
XX OS Homo sapiens.
XX
XX PN WO2004024097-A2.
XX
XX PD 25-MAR-2004.
XX
XX PF 15-SEP-2003; 2003WO-US029097.
XX
XX PR 16-SEP-2002; 2002US-0411392P.
XX
XX PA (GETH ) GENENTECH INC.
XX
XX PI Chiu H, Clark H, Dennis K, Pong S, Schoenfeld JR, Wood WI;
XX WU TD;
XX
XX DR WPI; 2004-329389/30.
XX
XX DR N-PSDB; ADL83101.
XX
XX New PRO polypeptide, useful for diagnosing and treating a B cell related
PT disorder, e.g. Burkitt's lymphoma, rheumatoid arthritis, autoimmune
PT mediated hemolytic anemia, myasthenia gravis or ankylosing spondylitis.
XX
XX Claim 10; Fig 304; 695pp; English.
XX
XX The present invention relates to PRO proteins and their coding sequences.
XX The PRO proteins are useful for diagnosing and treating a B cell related
XX disorder, e.g. X-linked infantile hypogammaglobulinemia, polysaccharide
XX antigen unresponsiveness, selective IgA deficiency, selective IgM
XX deficiency, selective deficiency of IgG subclasses, immunodeficiency with
XX hyper IgM, transient hypogammaglobulinemia of infancy, Burkitt's
XX lymphoma, intermediate lymphoma, follicular lymphoma, type II
XX hypersensitivity, rheumatoid arthritis, autoimmune mediated haemolytic
XX anaemia, myasthenia gravis, hypoadrenocorticism, glomerulonephritis, or
XX ankylosing spondylitis. The PRO proteins are also useful for preparing a
XX medicament for treating a condition that is responsive to the PRO
XX protein, e.g. cancer or immune-mediated inflammatory diseases. The PRO
XX coding sequences are useful as hybridization probes in chromosome and
XX gene mapping, in preparing PRO proteins, or in generating transgenic
XX animals or knockout animals, which in turn are useful in the development
XX and screening of therapeutically useful reagents.
XX
XX SQ Sequence 958 AA;
XX
XX Query Match 26.8%; Score 1753.5; DB 8; Length 958;
XX Best Local Similarity 49.0%; Pred. No. 1.5e-150;
XX Matches 341; Conservative 114; Mismatches 210; Indels 31; Gaps 9;
XX
QY 72 HPSLYFDAGETQAMRQKSRASHLHLFRAIRSAVTVMLSNPTTYLPPPKHADFAAKWNEY 131
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 43 HPMYFSAEVAELQLRASASHEHIAARLTAHVHTMLSSPLEYLPWDPKDYSAKWEIF 102
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 132 GNNLPPLALYCLCPEDKVAPEFVLEVMYDMRVGKWLVENAFGDEVPIGHSLTGATFAT 191
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 103 GNNLGALAMFCVLYPENIEARDMAKDYMERMAAQPSMLVKDAPWDEVLASHLVGFATAY 162
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 192 DFLYNLLDNHRRQKYLEKIWIITEEMYEYSKVRSGKQLLNHNQATNMIALTGALVTGV 251
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 163 DFLYNLSKTKQKEFLEVIANASGYMYETSYRRGWFQYLLHNHQPNTCWMALLTGSVL-M 221
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :

```

```
Qy 252 DKG--SKANIWKQAVDVMKMTFLNHHIVDGLDEGVAYGYSYTAQSVTQVFLAQRHFN 309
Db 222 NQGYLQEAAYLWTKQVLITMEKSLVLLREVDTGSLYEGVAYGYSYTRSLFQYMLVQRHFN 281
Qy 310 INNLNDNLWLMKHFYFATLLPGFQRTQVGIADSNYNMFYGPESQVFLDFKILKNGAGNW 369
Db 282 INHFCHPMLKQHPAFMYRTLLPGFQRTVAIADSNYNMFYGPESQVFLDFKILKNGAGNW 341
Qy 370 LAQQRKHPKDPMPVSTQARWSTLHTEYTWYDPQLTPQPPADYGTAKHTFENWGVVT 429
Db 342 LADQIRNRNVVEGTPSKQQRWCTLHTEFLWYDGLSKSVPPDPFGTTLHYFEDWGVVT 401
Qy 430 YGAGLPNTQNTFTVFSKGLGGRVAVDVIHFQPY-SWIDGWRSFNPGCHEPQDQNSTFA 488
Db 402 YGSALPAEINRSFLSFKSGLGGRVAVDVIHFQPY-KWIKGRNFMNAGHEPQDQNSTFA 461
Qy 489 PNGQVFSVSEALYGPKLHLNVLVAPSPSQCNKPNWEGQLGE--CAQWLKWTGSEVGDA 546
Db 462 PNGVFFITEALYGPKYTFNNVLMFSPVAVSKSCFSPWVGQVTEDCSSKWKYKHDLAASC 521
Qy 547 AGEIITASQHGEMVPSGEAVSASSAMRLKSVYRALLLNSOTLLVVDHIEROEDSPIN 606
Db 522 QGRVVAEEKNGVVFIRGEGVAGYNPQLNKNVQRNLLILHPQLLLVVDQHLGEEPSLE 581
Qy 607 SVSAFFHNLDDIDFKYIPYKFNRYNGAMVDVDAHYKMFDPDHGNSPMASIQEAQAAE 666
Db 582 TAASFFHNVDPFE---ETVVDGVHGAFIQRDGLYKMYWMDTGYSKATFASVTPYPRG 638
Qy 667 FKRWTFQVNVTFQWESITRIAYVFGPYINVSRCRPIIDSSNPGLOISLNNVNTHEVVS 726
Db 639 YPYNGTNYVNTVTHLRSPTIRAAVLFIGPSIDVQS-----FTVHGDSQQQLDVF 686
Qy 727 IVTDYHNLKTFPNVL-----GFGGFASV-ADQGOI 755
Db 687 IATSKHAYAT---YLTWTEATGQSAFAQVIADRHKI 719

RESULT 12
AAB93826
ID AAB93826 standard; protein; 314 AA.
XX AC AAB93826;
XX DT 26-JUN-2001 (first entry)
XX DE Human protein sequence SEQ ID NO:13638.
XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX OS Homo sapiens.
XX PN EP1074617-A2.
XX PD 07-FEB-2001.
XX PF 28-JUL-2000; 2000EP-00116126.
XX PR 29-JUL-1999; 99JP-00248036.
XX PR 27-AUG-1999; 99JP-00300253.
XX PR 11-JAN-2000; 2000JP-00118776.
XX PR 02-MAY-2000; 2000JP-00183767.
XX PR 09-JUN-2000; 2000JP-00241899.
XX PA (HELI-) HELIX RES INST.
XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX DR WPI; 2001-318749/34.
XX PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or
```

diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

Claim 8; SEQ ID NO 13638; 2537pp + Sequence Listing; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention

XX Sequence 314 AA;

Query Match 20.4%; Score 1331; DB 4; Length 314;
Best Local Similarity 32.4%; Pred. No. 1.3e-112;
Matches 313; Conservative 0; Mismatches 1; Indels 652; Gaps 1;

```
Qy 1 MPKGGAPPWIMALMFTCHLLFLALLMPAFSTFPEESVSNYSEAWFTDIDQFKTKQVDF 60
Db 1 MPKGGAPPWIMALMFTCHLLFLALLMPAFSTFPEESVSNYSEAWFTDIDQFKTKQVDF 60
Qy 61 RPNQKLKSMHPSLYFDAGEIQAMROKSRASHLHLPRAIRSATVTVMLSNPTYYLPPPKH 120
Db 61 RPNQKLKSMHPSLYFDAGEIQAMROKSRASHLHLPRAIRSATVTVMLSNPTYYLPPPKH 120
Qy 121 ADFAAKWNEIYGNLPPALYCLLPEDKVAPEFVLEMDRMVGYKDWLVENAPGDEVPI 180
Db 121 ADFAAKWNEIYGNLPPALYCLLPEDKVAPEFVLEMDRMVGYKDWLVENAPGDEVPI 180
Qy 181 GHSITGPATAFDFLYNLLDNHRRQKYLEKIWTVEEMVEYSKVRSGKQLLHNHQAATMI 240
Db 181 GHSITGPATAFDFLYN-----
Qy 241 ALLTGALVTGVKSGSKANIWKQAVVDVMEKTMFLNHHIVDGLDEGVAYGYSYTAQSVTQY 300
Db 197 -----
Qy 301 VFQAQRHFNINLNDNNLWLMKHFYFATLLPGFQRTQVGIADSNYNMFYGPESQVFLDFK 360
Db 197 -----
Qy 361 ILKNGAGNLAAQIRKHPKDPMPVSTQARWSTLHTEYTWYDPQLTPQPPADYGTAKIH 420
Db 197 -----
Qy 421 TFPNMGVVITYGAGLPNTQNTFTVFSKGLGGRVAVDVIHFQPYSWIDGWRSFNPGHEHP 480
Db 197 -----
Qy 481 DQNSFTFAPNGQVFSVSEALYGPKLHLNVLVAPSPSQCNKPNWEGQLGECQWLKWTG 540
Db 197 -----
Qy 541 EEVGDAAAGEIITASQHGEMVPSGEAVSASSAMRLKSVYRALLLNSOTLLVVDHIERQ 600
Db 197 -----
```

QY 601 EDSPINSVSAFFHNLIDFKYIPYKFNRYNGAMMDVDAHYKMFWDHGHGNSPMASIQE 660
Db 197 ----- 196
QY 661 AEQAAEFKRWTFQVNVTFQWESTITRIAYVYCPYINVSSCRFIDSSNPGLOISLVNN 720
Db 197 ----- 196
QY 721 TEHVSVITDYHNLKTRFNVLGRGGFASVADQGITRFGLGTOAIKVPVHDIRIIFPGF 780
Db 197 ----- 196
QY 781 KFNIAVGLILCISLVILTFQWRFLVSPKLRWLIIIVIALWFIELDVWSTCSQPICAK 840
Db 197 ----- 196
QY 841 WTRTEAGSKSLSEGHMDLPDVVITSLPGSGAEILKOLFNFSSDFLYIRVPTAYIDI 900
Db 197 -----CKKSLSEGHMDLPDVVITSLPGSGAEILKOLFNFSSDFLYIRVPTAYIDI 248
QY 901 PETELEIDSFVDACEWKVSDIRSGHFRLLRCWLQSLVODTKLHLQNIHLHEPNRGKLAQY 960
Db 249 PETELEIDSFVDACEWKVSDIRSGHFRLLRGWLQSLVODTKLHLQNIHLHEPNRGKLAQY 308
QY 961 FANWKD 966
Db 309 FANWKD 314
RESULT 13
ADRI0287
ID ADRI0287 standard; protein; 625 AA.
XX
AC ADRI0287;
XX
DT 04-NOV-2004 (first entry)
XX
DE Human protein useful for treating neurological disease Seq 3793.
XX
KW human; oligo-capping method; diagnostic marker; gene therapy;
KW osteoporosis; neurological disease; Alzheimer's disease;
KW Parkinson's disease; dementia; short memory; cancer;
KW sense or motor function; emotional reaction; fear response; panic;
KW osteopathic; neuroprotective; neurotropic; antiparkinsonian; cytosstatic;
KW Tranquilliser.
XX
OS Homo sapiens.
XX
PN EPI447413-A2.
XX
PD 18-AUG-2004.
XX
PF 12-FEB-2004; 2004EP-00003145.
XX
PR 14-FEB-2003; 2003JP-00102207.
PR 09-MAY-2003; 2003JP-00131452.
XX
XX (REAS-) RES ASSOC BIOTECHNOLOGY.
PA
XX Isogai T, Yamamoto J, Nishikawa T, Isono Y, Sugiyama T, Otsuki T;
PI Wakamatsu A, Ishii S, Nagai K, Irie R;
PI
XX WPI; 2004-583265/57.
DR N-PSDB; ADR08331.
XX
XX New 1995 cDNA, useful for treating osteoporosis, neurological diseases,
PT Alzheimer's diseases, Parkinson's disease, dementia and various cancers.
XX
XX Claim 1; SEQ ID NO 3793; 2686pp; English.
PS
XX This invention relates to novel, isolated full length human cDNA
XX molecules and the encoded proteins thereof. Specifically, it refers to
CC
CC

CC cDNA clones obtained by an oligo-capping method, where none of these
CC clones are identical to any known human mRNAs. The present invention
CC describes an immunoassay to identify agonists and antagonists, as well as
CC antibodies, antisense molecules and siRNAs that can all be used to bind
CC to and modulate expression of the cDNA molecules. As such, these
CC molecules are useful for diagnostic markers or therapeutic targets for
CC the various diseases or morbid states. In particular, they are useful in
CC gene therapy for treating osteoporosis, neurological disease, Alzheimer's
CC disease, Parkinson's disease, dementia, short memory and various cancers,
CC as well as for maintaining equilibrium of sense or motor function, and
CC for treating emotional reaction, fear response and panic. Accordingly,
CC they exhibit osteopathic, neuroprotective, neurotropic, antiparkinsonian,
CC cytosstatic and tranquilliser activities. This polypeptide is a protein
CC encoded by a full length human cDNA sequence of the invention. NOTE: This
CC sequence is not given in the sequence listing of the specification but
CC can be obtained on CD-ROM from the European Patent Office, Vienna Sub-
CC office.
XX
XX Sequence 625 AA;
SQ
Query Match 13.7%; Score 898; DB 8; Length 625;
Best Local Similarity 44.6%; Pred. No. 2e-72; Mismatches 131; Indels 28; Gaps 7;
Matches 180; Conservative 65;
QY 362 LKNGAGNLAQQIRKRPKDGMPVSTAQRMSTLHTEYIWDYDLPQTPQADYGTAKIHT 421
Db 1 MRNGSGNWLADQIRNRNVVEGPGTSGQRWCTLHTEFLWYDGLSKSVPPDPFGTPTLHY 60
QY 422 FPNWGVVYTGAGLPNTQNTNTVFSFKSGKLGRAVVDIVHFPQY-SWIDGWSRFPNGHEHP 480
Db 61 FEDWGVVYTGAGLPNTQNTNTVFSFKSGKLGRAVVDIVHFNKYNKDWIRNFGNAGHEHP 120
QY 481 DONSFTFAPNGQVFSVSEALYGPKLHLNNLVFAPSQQCNKPWEGOLGE--CAQWLKW 538
Db 121 DONSFTFAPNGVFPFTEALYGPKYTFNNVLMFSPAVSKCFSPVWGVQVTEDCSSKWSKY 180
QY 539 TGEYVGDAAGEIITASQHGEMVFSGEAVSAYSSAMRLKSVYRALLNLSQTLVVVDHIE 598
Db 181 KHDLAASCQGRVAAEEKNGVVFIRGEGVAYNPOLNKNVQNLILHLPOLLVLDQIH 240
QY 599 ROEDSPINSVSAFFHNLIDFKYIPYKFNRYNGAMMDVDAHYKMFWDHGHGNSPMASI 658
Db 241 LGEESPLETAASFHNVDVPFE--ETVVDGVHGAFIRQRDLGYKXVWMDDTGYSEKATF 297
QY 659 QEAQAAEFKRWTFQVNVTFQWESTITRIAYVYCPYINVSSCRFIDSSNPGLOISLVN 718
Db 298 ASVTYPRGYPNYNTNVNVTMHLRSPITRAAYLFGPSIDVQS-----FTVHG 345
QY 719 NNTHEWSIVTDYHNLKTRFNYL-----GFGGFASV-ADQOI 755
Db 346 DSQQLDVFIAATSKHAYT---YLTWGEATGQSFAQVIAADRHKI 386
RESULT 14
ADG27796
ID ADG27796 standard; protein; 128 AA.
XX
XX ADG27796;
XX
DT 26-FEB-2004 (first entry)
XX
DE Human novel protein amino acid sequence SeqID563.
XX
KW antiparkinsonian; haemostatic; neurotropic; neuroprotective; osteopathic;
KW anti-HIV; protozoacide; antifungal; immunosuppressive; antirheumatic;
KW antiarthritic; antidiabetic; antiallergic; antiinflammatory;
KW anticoagulant; cytosstatic; gene therapy; Parkinson's disease;
KW Alzheimer's disease; thrombocytopaenia; osteoporosis; osteoarthritis;
KW infection; HIV; Leishmania; malaria; fungal infection;
KW multiple sclerosis; rheumatoid arthritis;
KW insulin dependent diabetes mellitus; allergic reaction; food allergy;
KW insect allergy; allergic rhinitis; haemophilia; cancer; human.
XX

describes these compositions as useful for diagnosing, treating or preventing disorders such as cancer, haematopoietic diseases including anaemia and multiple myeloma, reproductive system disorders including prostaticitis and inguinal hernia, musculoskeletal diseases including systemic lupus erythematosus and gout, cardiovascular disease including arrhythmia and hypernatraemia, mixed fetal diseases including fetal alcohol syndrome and Down's syndrome, excretory diseases including urinary incontinence and renal disorders, neural or sensory disease including Alzheimer's disease and meningitis, respiratory disease including emphysema and occupational lung disease, endocrine diseases including diabetes and glomerulonephritis, digestive diseases including portal hypertension and irritable bowel syndrome and connective tissue or epithelial diseases including scleroderma and epidermolysis bullosa. As such, there are various activities such as cytostatic, antianemic, antiarthritic, antiasthmatic, anti-HIV immunosuppressive, antiinflammatory, antipsoriatic, antibacterial, osteopathic, dermatological, antigout, immunomodulator, antiarrhythmic, cardiant, neurotropic, antilipemic, nephrotropic, uropathic, neuroprotective, antiparkinsonian, tranquilizer, antidiabetic, anabolic, hypertensive and vulnervary. This polypeptide is a human secreted protein of the invention. NOTE: This sequence does not appear in the printed specification but has been obtained in electronic format from the US patent office at the following web site www.seqdata.uspto.gov/sequence.html; Document ID: 20040010132.

SQ Sequence 125 AA;

Query Match 8.7%; Score 568; DB 8; Length 125;
Best Local Similarity 100.0%; Pred. No. 2.5e-43;
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

| | | | |
|----|-----|---|-----|
| Qy | 811 | MRWILILVIALWFLIDVWSTCSQPTCAKWTTEAGSKKLSLSRGHMDLPDVVITSL | 870 |
| Db | 1 | MRWILILVIALWFLIDVWSTCSQPTCAKWTTEAGSKKLSLSRGHMDLPDVVITSL | 60 |
| Qy | 871 | PGSGAEILKQLFFNSSDFLYIRVPTAYIDIPETELEIDSFVDACEWK | 917 |
| Db | 61 | PGSGAEILKQLFFNSSDFLYIRVPTAYIDIPETELEIDSFVDACEWK | 107 |

Search completed: June 23, 2005, 08:43:30
Job time : 135.064 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2005, 08:34:49 ; Search time 46.0594 Seconds
(without alignments)
1980.512 Million cell updates/sec

Title: US-10-697-828-15

Perfect score: 6532

Sequence: 1 MPRGGAPPWIMALMFTGHL.....LIENICWTLMDRLGYPKFM 1222

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA.*
1: /cgn2_6/ptodata/1/iaa/5A COMB.pbp.*
2: /cgn2_6/ptodata/1/iaa/5B COMB.pbp.*
3: /cgn2_6/ptodata/1/iaa/6A COMB.pbp.*
4: /cgn2_6/ptodata/1/iaa/6B COMB.pbp.*
5: /cgn2_6/ptodata/1/iaa/PCFUS COMB.pbp.*
6: /cgn2_6/ptodata/1/iaa/backfilee1.pbp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------------|
| 1 | 568 | 8.7 | 126 | 4 | US-09-288-143-71 |
| 2 | 147 | 2.3 | 411 | 3 | US-09-015-188-2 |
| 3 | 145 | 2.2 | 2504 | 4 | US-08-851-567B-12 |
| 4 | 145 | 2.2 | 2504 | 4 | US-09-817-514A-8 |
| 5 | 135 | 2.1 | 1844 | 4 | US-08-851-567B-53 |
| 6 | 134 | 2.1 | 483 | 3 | US-09-263-023-2 |
| 7 | 134 | 2.1 | 483 | 4 | US-09-471-867-2 |
| 8 | 133 | 2.0 | 484 | 3 | US-09-263-023-4 |
| 9 | 133 | 2.0 | 484 | 4 | US-09-471-867-4 |
| 10 | 133 | 2.0 | 531 | 4 | US-09-949-016-6471 |
| 11 | 133 | 2.0 | 608 | 4 | US-09-949-016-9449 |
| 12 | 127 | 1.9 | 1456 | 4 | US-09-976-594-168 |
| 13 | 123.5 | 1.9 | 395 | 4 | US-09-949-016-7011 |
| 14 | 120.5 | 1.8 | 479 | 2 | US-08-899-514-2 |
| 15 | 119.5 | 1.8 | 390 | 4 | US-09-949-016-6813 |
| 16 | 119.5 | 1.8 | 431 | 4 | US-09-949-016-8893 |
| 17 | 115.5 | 1.8 | 386 | 4 | US-09-786-240-11 |
| 18 | 115 | 1.8 | 893 | 3 | US-09-514-302-4 |
| 19 | 115 | 1.8 | 893 | 4 | US-10-014-436-4 |
| 20 | 115 | 1.8 | 1938 | 3 | US-09-514-302-2 |
| 21 | 115 | 1.8 | 1938 | 4 | US-10-014-436-2 |
| 22 | 112 | 1.7 | 772 | 1 | US-08-258-639A-2 |
| 23 | 112 | 1.7 | 772 | 2 | US-08-900-951-2 |
| 24 | 112 | 1.7 | 772 | 5 | PCT-US95-07391A-2 |
| 25 | 111.5 | 1.7 | 462 | 2 | US-08-870-180B-13 |
| 26 | 111.5 | 1.7 | 462 | 3 | US-09-226-529-13 |
| 27 | 110.5 | 1.7 | 1455 | 3 | US-08-840-062-5 |

| | | | | | | |
|----|-------|-----|------|---|----------------------|-------------------|
| 28 | 108.5 | 1.7 | 386 | 3 | US-09-045-284A-2 | Sequence 2, Appli |
| 29 | 108.5 | 1.7 | 386 | 3 | US-09-190-911-1 | Sequence 1, Appli |
| 30 | 108 | 1.7 | 826 | 4 | US-09-252-991A-22143 | Sequence 22143, A |
| 31 | 105 | 1.6 | 391 | 4 | US-09-248-796A-15903 | Sequence 15903, A |
| 32 | 105 | 1.6 | 1180 | 3 | US-09-224-024-28 | Sequence 28, Appl |
| 33 | 105 | 1.6 | 1180 | 5 | PCT-US94-07902-28 | Sequence 28, Appl |
| 34 | 104.5 | 1.6 | 843 | 4 | US-09-248-796A-17210 | Sequence 17210, A |
| 35 | 104 | 1.6 | 1259 | 3 | US-09-134-001C-3757 | Sequence 3757, Ap |
| 36 | 104 | 1.6 | 1509 | 4 | US-09-676-519-27 | Sequence 27, Appl |
| 37 | 103.5 | 1.6 | 324 | 4 | US-09-270-767-43224 | Sequence 43224, A |
| 38 | 103 | 1.6 | 582 | 3 | US-09-091-725-17 | Sequence 17, Appl |
| 39 | 102 | 1.6 | 1227 | 4 | US-09-710-279-96 | Sequence 96, Appl |
| 40 | 101.5 | 1.6 | 594 | 3 | US-08-987-151-2 | Sequence 2, Appli |
| 41 | 101 | 1.5 | 574 | 4 | US-09-248-796A-18722 | Sequence 18722, A |
| 42 | 101 | 1.5 | 650 | 4 | US-09-248-796A-16387 | Sequence 16387, A |
| 43 | 100.5 | 1.5 | 1172 | 4 | US-09-328-352-6071 | Sequence 6071, Ap |
| 44 | 100 | 1.5 | 630 | 4 | US-09-538-092-384 | Sequence 384, App |
| 45 | 100 | 1.5 | 906 | 4 | US-09-417-039-11 | Sequence 11, Appl |

ALIGNMENTS

RESULT 1
US-09-288-143-71
; Sequence 71, Application US/09288143
; Patent No. 6433139
; GENERAL INFORMATION:
; APPLICANT: Brewer et al.
; TITLE OF INVENTION: 53 Human Secreted Proteins
; FILE REFERENCE: P2018P1
; CURRENT APPLICATION NUMBER: US/09/288,143
; CURRENT FILING DATE: 1999-04-08
; EARLIER APPLICATION NUMBER: PCT/US98/21142
; EARLIER FILING DATE: 1998-10-08
; EARLIER APPLICATION NUMBER: 60/061,463
; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: 60/061,529
; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: 60/071,498
; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: 60/061,527
; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: 60/061,536
; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: 60/061,532
; EARLIER FILING DATE: 1997-10-09
; NUMBER OF SEQ ID NOS: 219
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 71
; LENGTH: 126
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (126)
; OTHER INFORMATION: Xaa equals stop translation
; US-09-288-143-71

Query Match 8.7%; Score 568; DB 4; Length 126;
Best Local Similarity 100.0%; Pred. No. 3.8e-49;
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

| | | | |
|----|-----|--|-----|
| Qy | 811 | MRWILLIVIALWTELLDVWSTCSQPCAKWTEAGSKKLSSEGHMDLPDVTLSL | 870 |
| Db | 1 | MRWILLIVIALWTELLDVWSTCSQPCAKWTEAGSKKLSSEGHMDLPDVTLSL | 60 |
| Qy | 871 | PGSGAELTKQLFFNSSDFLYIRVPTAYIDIPETELEIDSFVDACEWK | 917 |
| Db | 61 | PGSGAELTKQLFFNSSDFLYIRVPTAYIDIPETELEIDSFVDACEWK | 107 |

RESULT 2

US-09-015-188-2
; Sequence 2, Application US/09015188C
; Patent No. 6399358
; GENERAL INFORMATION:
; APPLICANT: Williams, Kevin J
; APPLICANT: Tabas, Ira
; TITLE OF INVENTION: A Human Gene Encoding Human Chondroitin
; FILE REFERENCE: JEFF-0231
; CURRENT APPLICATION NUMBER: US/09/015,188C
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 411
; TYPE: PRN
; ORGANISM: Homo sapiens
US-09-015-188-2

Query Match 2.3%; Score 147; DB 3; Length 411;
Best Local Similarity 18.2%; Pred. No. 1.9e-05;
Matches 89; Conservative 58; Mismatches 149; Indels 194; Gaps 16;

Qy 813 WILVIALWFIEL-----LDVMSTC-----SOPICAKWTRTEAGSKSLSS 855
Db 5 WKAVLLALASIAIQTAIRTFKSFHTCPLAEAGLAERLC-----EESPTFAYNLSR 59
Qy 856 EGHMDLPDVVITSFGSGAEIKQLFFNSSDFLYTRVPTAYIDIPETELEDISFVDACE 915
Db 60 KTH-----ILILATRRSGSFGQLFNQHLDFVYLFEPL----- 93
Qy 916 WKVSDIRSGHFRLLRGWLQSLVQDTKLHLQNIHLHBPNGKLAQYFAMNKDKKFKFRE 975
Db 94 -----YHVNTLIPRTQG-----KSPADRV 115
Qy 976 SLPEQRSMKGAFDRDAEYI-----RALRRHLVYPSARPV-----LS 1013
Db 116 MCGASRDLRLSLVDCDLYFLENYKPPVNHVTDTRFRGASRVLCSPVCDPPGADLV 175
Qy 1014 LSSGSWTLKHFFQEVUGA-----SMRALYIVRD 1042
Db 176 LEEGDCVRKCGLLNLTVAAEACRSHVAIKTVRVPVNDLRALVDPRNLKVIQLVRD 235
Qy 1043 PRAWISMLYNSKPSLYSLKNVPEHLAKLFKIEGGKGNLNSGYAFEYELRKLKLSK 1102
Db 236 PRGILASRSTFRDY-----RLWRLWYGTGR-----KPYNLDTQL-TTVCEDF 279
Qy 1103 SNAVS--LLSHLWLANTAALRINTDLLPTSQYOLVKFEDIVHPPOKTTTERRIFAFLGIPL- 1159
Db 280 SNSVSTGLMRPPLKG-----KYMVLRYEDLARPNMKTEIYGFGLGIPLD 325
Qy 1160 -----SPASLNQILFATSNLFLVPLVEGEISPTNTNVKQNLPRDEIKLIENI 1207
Db 326 SHVARMTQNTNRGDTLGHKHYGTVRN-----SAATAEKWRFLSYDIVAFAQNA 375
Qy 1208 CWTLMDRLG 1217
Db 376 CQVLAQLGY 385

RESULT 3
US-08-851-567B-12
; Sequence 12, Application US/08851567B
; Patent No. 6528484
; GENERAL INFORMATION:
; APPLICANT: Ensign, Jerald C
; APPLICANT: Bowen, David J
; APPLICANT: Petell, James
; APPLICANT: Fatig, Raymond
; APPLICANT: Schoonover, Sue
; APPLICANT: Iffrench-Constant, Richard
; APPLICANT: Rocheleau, Thomas A.
; APPLICANT: Blackburn, Michael B.

APPLICANT: Hey, Timothy D.
APPLICANT: Merlo, Donald J.
APPLICANT: Orr, Gregory L.
APPLICANT: Roberts, Jean L.
APPLICANT: Strickland, James A.
APPLICANT: Guo, Lining
APPLICANT: Ciche, Todd A.
APPLICANT: Sukhapinda, Kitisri
TITLE OF INVENTION: Insecticidal Protein Toxins From Photorhabdus
NUMBER OF SEQUENCES: 88
CORRESPONDENCE ADDRESS:
ADDRESS: Dow Agrosciences Patent Department
STREET: 9330 Zionsville Road
CITY: Indianapolis
STATE: IN
COUNTRY: US
ZIP: 46268
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/851,567B
FILING DATE: 05-MAY-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/063,615
FILING DATE: 18-MAY-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/395,497
FILING DATE: 28-FEB-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/007,255
FILING DATE: 06-NOV-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/608,423
FILING DATE: 28-FEB-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/705,484
FILING DATE: 28-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 960296.93804
TELECOMMUNICATION INFORMATION:
TELEPHONE: 608-251-5000
TELEFAX: 608-251-9166
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 2504 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-851-567B-12

Query Match 2.2%; Score 145; DB 4; Length 2504;
Best Local Similarity 18.1%; Pred. No. 0.00091;
Matches 231; Conservative 180; Mismatches 443; Indels 422; Gaps 60;

Qy 29 FSTFEESVSNYSEWA-----VFTDDIDQFKTKQVDFRPNQKLKSLMHPSLYFD 78
Db 1003 FTDWERYNKRYSTWAGVSELVYYPENYVDPTQRIQGTQWMD-ALLQSIQSLQADNTYED 1061
Qy 79 AGEIQAMQKSRASHLHLFRAIRSAVTV-----MLSNPTYLPPPKHA-----D 122
Db 1062 AFKTY-LTSFEQVANLKVISAYHDNVNVDQGLTVFIGIDQAAPGTYYWRSVDHSCENGK 1120
Qy 123 FAA-----KWNIEYGNLPLALYCLLCPEDKVAFEVLEVMYKDKLVENAPGDEV 178
Db 1121 FAANAWGEWNKI-----TCAVNPWKNIIRPVV---YMSRL---YLLWLEQQSKSD- 1165
Qy 179 PIGSLTGFATAFDPLYNLNDHRRQKYLEKIW-----VITEEMYEYSKVRSMGKQLLH 232

```

Db 1166 -----DQKTTIYQ--YNLKLHRI--YDGMNTPTFDVTEKVNKNTSSTDAESL-- 1212
Qy 233 NHQATNMIALTGALVTGVDKSGKANIWKQAVVDVMEKTMFLNHIYDGLSDEGVAGSY 292
Db 1213 -----GLYCTGYQ-----EDTLVMFYMQSS-----YSSY 1239
Qy 293 T---AKSVTOYVFLAQRHFNINLND--NNLKMHFYFYATLLPGFQRTVGIADSNYNWF 347
Db 1240 TDNNAPVTGLYIFADMSDDMTNAQATNYWNSY-----PQFDTVMADPDS----- 1286
Qy 348 YGPESQLVFLDKFLKNGAGNWLAAQRIKRPKDGMPVSTPAQRWSTLHTIYIYDPLT 407
Db 1287 -----NKKVITRRVNNRYAEDVE-----IPSSV-----TSNSNYSWGDSHSLT 1323
Qy 408 POPPADYGTAKIHTFPNMGVVTYAGLPN-----TQNTFVS--FKSGKLGGRV 455
Db 1324 -----MLYGSVPNITFESAEDRLRSTNNALSIHNGVAGTRRI 1363
Qy 456 YDIVHFPYMWIDGRSFPNPGHEHPDQNSFTFAPNGQVVFSEALYGPKLSHLNNVLVPAP 515
Db 1364 QCNLMKQVYASLGDKFIYDSSFD--DANRFLNP-----LFGDENSDDSIYNE 1414
Qy 516 SPSSQCNKPWEGQGECAQMLKWTGE-----EVGDAAGEIITASQHGEMFVSGEAVSAYS 571
Db 1415 NPSSE--DKKWYFSSKDDNKTDYNGTQCIDAGTSNKDFYNNLQIEIEIVISVTGGWSSYK 1473
Qy 572 SAMRLKSVYRALLLNLSQTLVVDHIERQEDSPNSVSAFFHNLDDIDPKYIPYKFMRYN 631
Db 1474 IS-----NPKINI-----NTGIDSARKV--KVTVKAG 1496

Qy 632 GAMDVMDAHYKMFDFHHGNSPMASIQEABQAAEFKKRTQFVNVTFQEST--ITRIA 689
Db 1497 G-----DQIFADNSTYVQ-----QAPSPFEMIYQFNLTIDCKNLAFIDNOA 1542
Qy 690 YVFYGPYINVSRCRFDSSNPGLOISLNVNNTHEVSVITDYHNL-----KTRFNYL 741
Db 1543 HIEIDFTAQDGRFLGAEFTIIPVTKVGLGTENVIALYSENNGVQVMOIGAYRTRNL 1602
Qy 742 GFGFASVADOG-----QITRFGLTQAIKVPVRHDLRIFFPGFKFIAGVLI 789
Db 1603 FAQQLVSRANRGIDAVLSMETQNTQEPQLGAGTVVQLVLDKYDESIGHTNKSFAIEYVDI 1662
Qy 790 LCISAVILTQWRFPYLSFRKLMRWILLVIALWPIELL-----DVMSTCSQPICAKW----- 841
Db 1663 FKENDSFVIYQELSETSQTVK-----VFLSYPIEATGNKNLW-----VRKYQKET 1711
Qy 842 -----TRTEAS-----GSKKLSSEGHMDLPDVVVITSLPGSGA--- 875
Db 1712 TDKILFDRDEKPHGWFLSDDHKTFSGLSAQALKNDSEPMDF-----SGANAL 1761
Qy 876 -----EILKQLFFNSD--FLYIRVPTAYIIDIPELEIDSFVDAECWKV 918
Db 1762 YFWEFLYTPMMAHRLLEQNFDAANHFVYMSPSGYI-----VDGKIAIYHMNV 1813
Qy 919 SDIRSGHFRLLRGWLQSLVODTKLHLQNLHLHEPN-----RGKLAQYFA-----MN 964
Db 1814 -----RPLEEDTSMNAQQLDSTDPDAVAQDDPMHYKVATFMATDILLMAR 1858
Qy 965 KDKKKFKRRRSLPQRSQMKGAFDRDAEYIRALR-----RHLVYPSARPVLSSSGS 1018
Db 1859 GDAAYRQLERDTLAEAKW-----YTOALNLLGDEPQVMLSTTWANPTLGNAAK 1908
Qy 1019 WTLKHLFFQEVUGASMRALYIVRDDPRAMYMLYNSKPSLSYLSKXNVPHEHLAKLFIEGK 1078
Db 1909 TQOQVR--QQVL--TQRLNSRVKTP-----LLGTANSLAL--FLPOENSKL-----K 1951
Qy 1079 G-----KCNLSGYAFEPYPRKELSKSNVAVLSLHMLANTAAALRINTDLLPT 1130
Db 1952 GWRTLAORMFNRLNLSIDGQPLSLPLYAKPADPKALLSAVSAQSGA----- 2001
Qy 1131 SYQLVKFEDIVH--FPQ 1145

```

```

Db 2002 --DLPKAPLTIHRFPQ 2015

RESULT 4
US-09-817-514A-8
; Sequence 8, Application US/09817514A
; Patent No. 6639129
; GENERAL INFORMATION:
; APPLICANT: ffrench-Constant, Richard
; APPLICANT: Bowen, David
; APPLICANT: Rocheleau, Thomas
; APPLICANT: Waterfield, Nicholas
; TITLE OF INVENTION: DNA SEQUENCES FROM PHOTORHABDUS LUMINESCENS
; FILE REFERENCE: 61645
; CURRENT APPLICATION NUMBER: US/09/817.514A
; PRIOR FILING DATE: 2000-03-26
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 2504
; TYPE: PR
; ORGANISM: Photorhabdus luminescens
US-09-817-514A-8

Query Match 2.2% Score 145; DB 4; Length 2504;
Best Local Similarity 18.1% Pred. No. 0.00091;
Matches 231; Conservative 180; Mismatches 443; Indels 422; Gaps 60;

Qy 29 FSTFESVSNSYSEWA-----VFTDIDQPKTKQVQDFRPNQKLKKSMLHPSLYFD 78
Db 1003 FTDWERYNKYSTWAGVSELVYYPENYVDPQRIQOTKMMD-ALLQSIQSLNADTVED 1061
Qy 79 AGELOAMQKRSRASHLHLFRAIRSAVTV-----MLSNPTLYLPPPKHA-----D 122
Db 1062 AFKTY-LTSPQVANLKVISAYHDNVNVDQGLTVFIGIDQAAPGYVWRSVDHSCENGK 1120
Qy 123 FAA-----KMEIYGNLPLALYCLCPEDKVAPEFVLEVMYDMRVGKDWLVENAPGDEV 178
Db 1121 FAANAWGWNKI-----TCVNPWKNIIRPVV--YMSRL--YLLWLEQSKSD- 1165
Qy 179 PIGHSLTGFPATFDPLYNLLDNHRRQKYLEKIW-----VITEWYKSVKRSCKQLLH 232
Db 1166 -----DGKTTIYQ--YNLKLHRI--YDGMNTPTFDVTEKVNKNTSSTDAESL-- 1212
Qy 233 NHQATNMIALTGALVTGVDKSGKANIWKQAVVDVMEKTMFLNHIYDGLSDEGVAGSY 292
Db 1213 -----GLYCTGYQ-----EDTLVMFYMQSS-----YSSY 1239
Qy 293 T---AKSVTOYVFLAQRHFNINLND--NNLKMHFYFYATLLPGFQRTVGIADSNYNWF 347
Db 1240 TDNNAPVTGLYIFADMSDDMTNAQATNYWNSY-----PQFDTVMADPDS----- 1286
Qy 348 YGPESQLVFLDKFLKNGAGNWLAAQRIKRPKDGMPVSTPAQRWSTLHTIYIYDPLT 407
Db 1287 -----NKKVITRRVNNRYAEDVE-----IPSSV-----TSNSNYSWGDSHSLT 1323
Qy 408 POPPADYGTAKIHTFPNMGVVTYAGLPN-----TQNTFVS--FKSGKLGGRV 455
Db 1324 -----MLYGSVPNITFESAEDRLRSTNNALSIHNGVAGTRRI 1363
Qy 456 YDIVHFPYMWIDGRSFPNPGHEHPDQNSFTFAPNGQVVFSEALYGPKLSHLNNVLVPAP 515
Db 1364 QCNLMKQVYASLGDKFIYDSSFD--DANRFLNP-----LFGDENSDDSIYNE 1414
Qy 516 SPSSQCNKPWEGQGECAQMLKWTGE-----EVGDAAGEIITASQHGEMFVSGEAVSAYS 571
Db 1415 NPSSE--DKKWYFSSKDDNKTDYNGTQCIDAGTSNKDFYNNLQIEIEIVISVTGGWSSYK 1473
Qy 572 SAMRLKSVYRALLLNLSQTLVVDHIERQEDSPNSVSAFFHNLDDIDPKYIPYKFMRYN 631
Db 1474 IS-----NPKINI-----NTGIDSARKV--KVTVKAG 1496

```

QY 632 GAMDVMDAHYKMFWDHGHGNSPMASIQEAAEFKKRWTFQVNVTFQWEST--ITRIA 689
Db 1497 G-----DDQFTADNSYVFO-----QAPSPSEMIYQFNLTIDCKNLNFDNQ 1542
QY 690 YVFGPYINSSCRFDSSPGIQLISLVNNTTHVSVITDYHNL-----KTRFNYL 741
Db 1543 HIEIDFTATAQDGRFLGAETFIIPVTKKVLGTENVIALYSENNGVQVMOIGAVRTRLNTL 1602
QY 742 GFGFASVADQG-----QITRFLGLGTQAIKVPVRHRIIIPPGFKFNIAVGLI 789
Db 1603 FAQQLVSRANRGIDAVLSMETQNIQEPQLGAGTYVQLVLDKYDESHGNTKNSFAIBYVDI 1662
QY 790 LCTSLVLTITQWRFLYFRKLMWILITVIALFIEL-----DWMSTCSQPICAKW--- 841
Db 1663 FKENDSVIYQGLSETSQTVK-----VFLSFIEATGNKWLH-----VRAKTOKET 1711
QY 842 -----TRTBAE-----GSKSLSEGHMHMDLPDVVITSLPGSGA--- 875
Db 1712 TDKILPDRDEKDPHGWFSLDDHKTFSGLSSAQALKNDSEPMDF-----SGANAL 1761
QY 876 -----ELKQLFNSSD--FLYIRVPTAYIDIPETLEIDSFVDACEWKV 918
Db 1762 YFWELFYTPMMAHRLLOBNFQDAANHWPYVWSPSGYI-----VDGKIATYHNV 1813
QY 919 SDIRSGHFRLLRGWLQSLVODTKLHLQNIHLHPN-----RGKLAOVFA-----MN 964
Db 1814 -----RLEEDTSNAQQLDSTDPDAVQDDPMHYKVATFMATLDLLMAR 1858
QY 965 KDKRKRFRRESPEQRSMKGAFDRDAEYIRALR-----RHLVYPSARPVLSSGS 1018
Db 1859 GDAAAYRQLERDTLAEAKW-----YTQALNLLGDPEQVWLSTTWNAPTIGNAASK 1908
QY 1019 WTLKLHFQEVILGASMRALYIVDRPRAWIYMSLYNSKPSLYSLKNVPEHLAKLFKEGCK 1078
Db 1909 TTOQVR--QQVL-TQLRLSRVKT-----LLGTANSLSLAL-FLPQENSKL-----K 1951
QY 1079 G-----KCNLSNGVAFEYEPRLKRLSKSNVAVSLLSHLWLANTAAALRINTDLLPT 1130
Db 1952 GYWRTLAQRNFWNLHNSIDQPLSLPYAKPADPKALUSAAVSASQGG----- 2001
QY 1131 SYQLVKFEDIVH-FPQ 1145
Db 2002 --DLPKAPLIHRFPQ 2015
RESULT 5
US-08-851-567B-53
Sequence 53, Application US/08851567B
Patent No. 6528484
GENERAL INFORMATION:
APPLICANT: Ensign, Jerald C
APPLICANT: Bowen, David J
APPLICANT: Petell, James
APPLICANT: Fatig, Raymond
APPLICANT: Schoonover, Sue
APPLICANT: ffrrench-Constant, Richard
APPLICANT: Rocheleau, Thomas A.
APPLICANT: Blackburn, Michael B.
APPLICANT: Hey, Timothy D.
APPLICANT: Merlo, Donald J.
APPLICANT: Orr, Gregory L.
APPLICANT: Roberts, Jean L.
APPLICANT: Strickland, James A.
APPLICANT: Guo, Lining
APPLICANT: Ciche, Todd A.
APPLICANT: Sukhapinda, Kitieri
TITLE OF INVENTION: Insecticidal Protein Toxins From Photorhabdus
NUMBER OF SEQUENCES: 88
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dow AgroSciences Patent Department
STREET: 9330 Zionsville Road
CITY: Indianapolis

STATE: IN
COUNTRY: US
ZIP: 46268
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/851,567B
FILING DATE: 05-MAY-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/063,615
FILING DATE: 18-MAY-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/395,497
FILING DATE: 28-FEB-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/007,255
FILING DATE: 06-NOV-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/608,423
FILING DATE: 28-FEB-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/705,484
FILING DATE: 28-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 960296.93804
TELECOMMUNICATION INFORMATION:
TELEPHONE: 608-251-5000
TELEFAX: 608-251-9166
INFORMATION FOR SEQ ID NO: 53:
SEQUENCE CHARACTERISTICS:
LENGTH: 1844 amino acids
TYPE: amino acids
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-851-567B-53
Query Match 2.1%; Score 135; DB 4; Length 1844;
Best Local Similarity 17.6%; Pred No. 0.0053;
Matches 186; Conservative 150; Mismatches 365; Indels 354; Gaps 48;
QY 29 FSTFEESVSNYSEWA-----VFTDDIDQFKTKQVQDFRPNQKLKSLMLPSLYFD 78
Db 916 FTDWERYNKRYSTWAGVSELVYYPENYVDPTQIGQTKMD-ALLQSIHQSLNADTVED 974
QY 79 AGEIOAMRQKSPASHLHLFRAIRSAVTV-----MLSNPTYLPPPKGHA-----D 122
Db 975 AFKTY-LTSFEQVANLKVISAYHDNVNDQGLTYFIGIDQAAPGTYVWRSVDHSCENK 1033
QY 123 FAA-----KWNEIYGNLPLALYCLLCPEDKVAPEFVLEVMYDMVGYKDWLVENAPGDEV 178
Db 1034 FAANAWGEWNI-----TCAVNFWKNIIRPVV--YMSRL--YLLMLEQQSKSD- 1078
QY 179 PIGHSLTGFATAFDFLYNLNHRHQKYLEKIW-----VITEEMEYYSKVRSWKQLLH 232
Db 1079 -----DGKTTIYQ--YNLKLAIH--YDGSWNTPTFTDVTETKVNTTSSTDAESL-- 1125
QY 233 NHQATNMIALTLGALVTGDKGSKANIWKQAVVDVWMEKTMELLNHIVDGSLDEGVAYGSY 292
Db 1126 -----GLYCTGYQG-----EDTLVVMFYMOSS-----YSSY 1152
QY 293 T---AKSVTQVFLAQRHFNINNLID--NNWLKMHFWFYATILLPGFQRTVGIADSNYNWF 347
Db 1153 TDNNAPVTGLYIFADMSSDNMTNAQATNYWNSY-----PQFDIVMADPDS----- 1199
QY 348 YGPESQLVFLDKFILKNGAGNWLAAQOIRKIRPKGPMVPSTAQRWSTLHTYIWTDPOLT 407
Db 1200 -----NKKVITRRVNRNRYAEDYE-----IPSSV-----TSNSNYSWGDHSLT 1236

```

QY 408 PQPADYGTAKIHTFPNKGVVYVYAGLPN-----TOTNTFVS-FKSGKLGRAV 455
Db 1237 -----MLYGGSPNITFESAEDLRLSTNMLSIHHNGVAGTRRI 1276

QY 456 YDIVHFOYGNIDGWRSPNFGHEPDQNSFTFAPNGQVFVSEALYKPLSHLNNVLVFP 515
Db 1277 QCNLMKQVSLGDKFIYDSSFD--DANRNLV----LF-----KFGKDNSSDDSIICYNE 1327

QY 516 SPSSQCNKPEGQGBCAQWLKWTGE-----EVGDAAGEIITASQHGEMVFSGEAVSAYS 571
Db 1328 NPSSE-DKKWYFSSKDDNKTDYNGGTCIDAGTSNKDFYVNLQIEIVISVTGGYWSYK 1386

QY 572 SAMRLKSVYRALLLNSQTLVVDHIERQBDSPNSVSAPFNHLDIDFKIPIPKFMRNRYN 631
Db 1387 IS-----NPNIN-----NTGIDSARV-KVTYKAG 1409

QY 632 GAMDMVDAHYKMFWDHNGSPNASIOEAAEFKRWTFQVNVTFQWEST--ITRIA 689
Db 1410 G-----DDQIFADNSTVPO-----QPAPSPFEMIYQFNLTIDCKNLNFIDNOA 1455

QY 690 YVFGPVYINSSCRFDISSNPGLOISLNVNNTHEVSVITVDYHNL-----KTRENYL 741
Db 1456 HIEIDFTATAQDGRFLGAETFIIPVTKVLGTENVIALYSENNGVQMQGAVKTRNLTL 1515

QY 742 GFGGFASVADQ-----QITRFLGLTQAIKVPVRHDIRIIPFGFKFNIAVGLI 789
Db 1516 FAQOLVSRANGIDAVLSMETONIQEQLGNGTVQLVLDKYDSIHGNTKNSFAIEYVDI 1575

QY 790 LCISLVILTQWRFYLSFRKLMRWILVIALMFIELL-----DYMSTCSQPICAKW---- 841
Db 1576 FKENDSFVIYQGLSETSTQTVK-----VFLSYFIEATGNKNHLW-----VRAKYQKET 1624

QY 842 -----TRTEAE-----GSKKLSSEGHMDLPDVVITSLPGSGA--- 875
Db 1625 TDKILFRTDEKPHGFWLSDHKTFTGLSQAQALKNDSSEMPD-----SGANAL 1674

QY 876 -----EILKQLFNSPD--FLYIRVPTAYIDIPELEIDSFVDACEWKV 918
Db 1675 YFWELFYTPMMAHRLGLQNFQDAANHFYVWSPSGYI-----VDGKIAIYHNW 1726

QY 919 SDIRSGHFRLLRGLQSLVQDTKLHLQNIHLHFN 953
Db 1727 -----RPLEEDTSNAQQLDSTDPD 1746

RESULT 6
US-09-263-023-2
; Sequence 2, Application US/09263023
; Patent No. 6037159
; GENERAL INFORMATION:
; APPLICANT: Uchimura, Kenji
; APPLICANT: Muramatsu, Hideki
; APPLICANT: Kadomatsu, Kenji
; APPLICANT: Kannagi, Reiji
; APPLICANT: Habuchi, Osami
; APPLICANT: Muramatsu, Takashi
; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-SULFOTRANSFERASE AND
; FILE REFERENCE: TOYAMA1.001AUS
; CURRENT APPLICATION NUMBER: US/09/263,023
; EARLIER FILING DATE: 1999-03-05
; EARLIER FILING DATE: 1998-03-05
; EARLIER FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 483
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-263-023-2

QY 408 PQPADYGTAKIHTFPNKGVVYVYAGLPN-----TOTNTFVS-FKSGKLGRAV 455
Db 1237 -----MLYGGSPNITFESAEDLRLSTNMLSIHHNGVAGTRRI 1276

QY 456 YDIVHFOYGNIDGWRSPNFGHEPDQNSFTFAPNGQVFVSEALYKPLSHLNNVLVFP 515
Db 1277 QCNLMKQVSLGDKFIYDSSFD--DANRNLV----LF-----KFGKDNSSDDSIICYNE 1327

QY 516 SPSSQCNKPEGQGBCAQWLKWTGE-----EVGDAAGEIITASQHGEMVFSGEAVSAYS 571
Db 1328 NPSSE-DKKWYFSSKDDNKTDYNGGTCIDAGTSNKDFYVNLQIEIVISVTGGYWSYK 1386

QY 572 SAMRLKSVYRALLLNSQTLVVDHIERQBDSPNSVSAPFNHLDIDFKIPIPKFMRNRYN 631
Db 1387 IS-----NPNIN-----NTGIDSARV-KVTYKAG 1409

QY 632 GAMDMVDAHYKMFWDHNGSPNASIOEAAEFKRWTFQVNVTFQWEST--ITRIA 689
Db 1410 G-----DDQIFADNSTVPO-----QPAPSPFEMIYQFNLTIDCKNLNFIDNOA 1455

QY 690 YVFGPVYINSSCRFDISSNPGLOISLNVNNTHEVSVITVDYHNL-----KTRENYL 741
Db 1456 HIEIDFTATAQDGRFLGAETFIIPVTKVLGTENVIALYSENNGVQMQGAVKTRNLTL 1515

QY 742 GFGGFASVADQ-----QITRFLGLTQAIKVPVRHDIRIIPFGFKFNIAVGLI 789
Db 1516 FAQOLVSRANGIDAVLSMETONIQEQLGNGTVQLVLDKYDSIHGNTKNSFAIEYVDI 1575

QY 790 LCISLVILTQWRFYLSFRKLMRWILVIALMFIELL-----DYMSTCSQPICAKW---- 841
Db 1576 FKENDSFVIYQGLSETSTQTVK-----VFLSYFIEATGNKNHLW-----VRAKYQKET 1624

QY 842 -----TRTEAE-----GSKKLSSEGHMDLPDVVITSLPGSGA--- 875
Db 1625 TDKILFRTDEKPHGFWLSDHKTFTGLSQAQALKNDSSEMPD-----SGANAL 1674

QY 876 -----EILKQLFNSPD--FLYIRVPTAYIDIPELEIDSFVDACEWKV 918
Db 1675 YFWELFYTPMMAHRLGLQNFQDAANHFYVWSPSGYI-----VDGKIAIYHNW 1726

QY 919 SDIRSGHFRLLRGLQSLVQDTKLHLQNIHLHFN 953
Db 1727 -----RPLEEDTSNAQQLDSTDPD 1746

RESULT 6
US-09-263-023-2
; Sequence 2, Application US/09263023
; Patent No. 6037159
; GENERAL INFORMATION:
; APPLICANT: Uchimura, Kenji
; APPLICANT: Muramatsu, Hideki
; APPLICANT: Kadomatsu, Kenji
; APPLICANT: Kannagi, Reiji
; APPLICANT: Habuchi, Osami
; APPLICANT: Muramatsu, Takashi
; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-SULFOTRANSFERASE AND
; FILE REFERENCE: TOYAMA1.001AUS
; CURRENT APPLICATION NUMBER: US/09/263,023
; EARLIER FILING DATE: 1999-03-05
; EARLIER FILING DATE: 1998-03-05
; EARLIER FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 483
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-263-023-2

```

```

Query Match 2.1%; Score 134; DB 3; Length 483;
Best Local Similarity 26.2%; Pred. No. 0.00054;
Matches 53; Conservative 36; Mismatches 81; Indels 32; Gaps 9;

QY 1034 MRALYIVRDPRAWIYMLYNSKPSLY--SLKNVPE-----HLAKLFKIEGK-GKCNLNS 1085
Db 278 LKVIHLVRDPRV--VASSRIRSRHGLIRSLQVVRSDPRHRMPFLEAAGHLGAKKEGM 336

QY 1086 GYAFEPYELRKLKSKSNVSLSHLWLAANTAAALRINTDLLPTSQVLKVFEDIVHPQ 1145
Db 337 GGPADYHAL-----GAMEVICNSMAKTLOTALQ--PPDWLQGHVLYVRYEDLVGDPV 386

QY 1146 KTERIFAFIGIPLSPA-----SLNQILFATSTNLFYLPYEGEISPTN-----TNVWKQNL 1197
Db 387 KTLRRVYDFVGLLVSPMEQFALNMTSGSGSSSKPFV-----VSARNATQAANAARWALT 441

QY 1198 RDEIKLIENICWTLMRLGYPK 1219
Db 442 FQIKQVEEFCYQPMVILGYER 463

RESULT 7
US-09-471-867-2
; Sequence 2, Application US/09471867
; Patent No. 6455289
; GENERAL INFORMATION:
; APPLICANT: Uchimura, Kenji
; APPLICANT: Muramatsu, Hideki
; APPLICANT: Kadomatsu, Kenji
; APPLICANT: Kannagi, Reiji
; APPLICANT: Habuchi, Osami
; APPLICANT: Muramatsu, Takashi
; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-SULFOTRANSFERASE AND
; FILE REFERENCE: TOYAMA1.001AUS
; CURRENT APPLICATION NUMBER: US/09/471,867
; CURRENT FILING DATE: 1999-12-23
; EARLIER FILING DATE: 1999-03-05
; EARLIER FILING DATE: 1999-03-05
; EARLIER FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 483
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-471-867-2

Query Match 2.1%; Score 134; DB 4; Length 483;
Best Local Similarity 26.2%; Pred. No. 0.00054;
Matches 53; Conservative 36; Mismatches 81; Indels 32; Gaps 9;

QY 1034 MRALYIVRDPRAWIYMLYNSKPSLY--SLKNVPE-----HLAKLFKIEGK-GKCNLNS 1085
Db 278 LKVIHLVRDPRV--VASSRIRSRHGLIRSLQVVRSDPRHRMPFLEAAGHLGAKKEGM 336

QY 1086 GYAFEPYELRKLKSKSNVSLSHLWLAANTAAALRINTDLLPTSQVLKVFEDIVHPQ 1145
Db 337 GGPADYHAL-----GAMEVICNSMAKTLOTALQ--PPDWLQGHVLYVRYEDLVGDPV 386

QY 1146 KTERIFAFIGIPLSPA-----SLNQILFATSTNLFYLPYEGEISPTN-----TNVWKQNL 1197
Db 387 KTLRRVYDFVGLLVSPMEQFALNMTSGSGSSSKPFV-----VSARNATQAANAARWALT 441

QY 1198 RDEIKLIENICWTLMRLGYPK 1219
Db 442 FQIKQVEEFCYQPMVILGYER 463

```


RESULT 11

US-09-949-016-9449
; Sequence 9449, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9449
; LENGTH: 608
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-9449

Query Match 2.0%; Score 133; DB 4; Length 608;
Best Local Similarity 26.2%; Pred. No. 0.0011;
Matches 53; Conservative 36; Mismatches 81; Indels 32; Gaps 9;

Qy 1034 MRALYIVDRPRAWYMLYNSKPSLY--SLKNVPE-----HLAKLFKEGK-KCKNLNS 1085
Db 403 LKVIHLVRDPRA-VASSKIRSRHGLIRESLOVRSRDPRAHRMPFLEAAGHLGAKKEGV 461

Qy 1086 GYAEYEPRLKELSKSNVSLSHLWLANAALRINTLLTSVLVKFEDIVHFPQ 1145
Db 462 GGPADYHAL-----GAMEVICNSMAKTLQALQ-PPDWLOGLYLWRYEDLVGDPV 511

Qy 1146 KTERIFAFILGIPSPA-----SLNQILFATSTNLFYPYEGEIGPTN-----TNVWKQNL 1197
Db 512 KTLRRVDFVGLLVSPMEQFALNWTSGSGSSKPFV-----VSARNATQANAWRTALT 566

Qy 1198 RDEIKLIENICWTLMRLGYPK 1219
Db 567 FQIKQVEFCYQPMVILGYER 588

RESULT 12

US-09-976-594-168
; Sequence 168, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 168
; LENGTH: 1456
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 1674368CD1
US-09-976-594-168

Query Match 1.9%; Score 127; DB 4; Length 1456;
Best Local Similarity 18.7%; Pred. No. 0.022;
Matches 138; Conservative 97; Mismatches 284; Indels 218; Gaps 36;

Qy 180 IGHSLTGFATAFDFLYN-----LLDNHRRQKYLEKIWVITEEMYEY---SKVRSWG--- 227
Db 518 IGHSLTGFATAFDFLYN-----LLDNHRRQKYLEKIWVITEEMYEY---SKVRSWG--- 227
Qy 228 -----KQLLNHNQATNMIALLTG--ALVTGVYDKSGSKANIKQAVVDVMEKTMFLNLHIVDG 281
Db 578 WTIEEVRFTHMNSDMPGRKPGCVAMRTGIAGG-----LWD--VLKCKEAKFVCKHWAEG 631
Qy 282 -----SLDEGVAYGYSYAKSVTQVYVFLAQRHF-----NIN 311
Db 632 VTHPPKPTTTPKPCPEDWGCASRTSLCFKLYAKGHEKKTWPEBRSDFCRALGGDLASIN 691
Qy 312 NLNN-----WL-----KQH--FW--FYATLLPGFQRTVGIADSNYNWFYGPBSOLVFLD 358
Db 692 NKEEQOTIWLITASGSYHKLFWLGLTYGSPSEGTWSDGSPVSYENWAYGEPNNYQNV 751
Qy 359 -----KFIKNGAGNWLAAQIIRKIR-----PKDGMVSTAGRWSTLH 396
Db 752 YGELKGDPTMSWINDINCEHLNNMIC-QIQKGTPEPTPAQDNP--PVTEDGW-VIY 807
Qy 397 TEYIWDPPQLTPOPPADYGTAKIHTFPNMGVVTYAGLNPQTNTFTVFSKSGKLGSRVY 456
Db 808 KDYY-----FSKEKTMNARAFCKRNFGL-----VSIQSESEKKFLWKYVNRNDAQSAY 860
Qy 457 DIVHF-----QPYSWIDG-----WRSFNPCHHPDQNSFTFAPNG-----QVF 494
Db 861 FIGLLISLDDKFFAWMDGSKVDYVSWATGEPNFANEDENCVTMYNSGFWNDINGYPN 920
Qy 495 VSEALYGPKLHLN--NLVFPAPSPSQCKNPWEGQLGECANWLKWTGEE----- 542
Db 921 ICOR-----HNSINATTMPMTSPVSGCKEGMNFYSNKCFKIFGFMBEERKNQOEARKA 976
Qy 543 -VGDAAGEIITASQHGMMVFSGEAVSAYSSAMRLKSVYRALLLNSTLLVVDHIERQE 601
Db 977 CIGFG-----GNLVSIOEKEQAFITYHMKDSTFSAWTGLND----- 1013
Qy 602 DSPINSVSFAFFHNLIDFKIPIYKFMNR--YNG---AMDMVMDAHYKMF-----WF 647
Db 1014 --VNSEHTF--LWTDGRGVHYTNMGKYPGGRSSLSYEDADCVCVIIGCASNEAGKWM 1067
Qy 648 DHGNS-----PMASIOEABQAAEFKRWTFQVNVTFQMESTIT----- 686
Db 1068 DDTCDSKRGYICOTRSDPSLTNPATIQ--TDGFVKYKSSYSMLMRQKQFQWHEAETCYK 1126
Qy 687 --RIAYVFGYINVSSECFIDSSNPGLOISLVNNTHEH-----VVSIVTDVHN 733
Db 1127 NSLIASIL-DPYSNAPAWLQMETSNERNVIALNSLTDNQYTTWTDKWRVRYTNWAADPEK 1185
Qy 734 LKTRFNYLGFPGFASVA 750
Db 1186 LKSACVYLDLDGYWKA 1202

RESULT 13

US-09-949-016-7011
; Sequence 7011, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0

```
; SEQ ID NO 7011
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Human
; US-09-949-016-7011

Query Match      1.9%; Score 123.5; DB 4; Length 395;
Best Local Similarity 23.3%; Pred. No. 0.0043;
Matches 90; Conservative 53; Mismatches 140; Indels 103; Gaps 20;

Qy 865 VVITSLPGSABILKQLFFNSSDFLYIRVPTAYIDIPETELEIDSFVDACEWKV-SDIRS 923
Db 43 VLVLSSWRSGSFFVQLFNQHPDVFYLMPEA-----WHVWTTLSQ 82
Qy 924 GHFELLRGWLQSLVQDTKL-HLQNIHLHEPNRGKLAQYFAMNKDKKFKKFRRESLPEQRS 982
Db 83 GSAATLHMAVRDLVRSVFLCDMDVFDAYLPWRRNLSDLF-----QWAVSRALCSPAC 136
Qy 983 QMKGAFORDAEYTRALARHLVYTPS---ARPVLSSSGSWTLKLHPFO-EVL-----G 1031
Db 137 ----AFPRGAISSAECVKPCARQSFTLAREACRSYSHVVLEKVRFNLOVLPILLS 192
Qy 1032 ASMRALYIVDRPRAWIYMLYNKPSLYSLKNVPEHLAKLFKIEGG--KGKCNLNSGYAF 1089
Db 193 LNLRIVHLVRDPRAVLSR-----EOTAKALARDNGIVLG-----TNGTWV 233
Qy 1090 EYEP-LR--KELSKSKSNVSLLSHLWLANTAALRINTDLLTSYQLVKFEDIVHFPQK 1146
Db 234 EADPGLRVREVCR-----SHVRIAE-AATLK-PPFPLRGRVLRVFEFLAREPLA 282
Qy 1147 TTERIFAFLGIPISP---ASINQI-----LFATSTNLFYLPYEGEISPTN-TNV 1191
Db 283 EIRALYAFTGLSLTPQLEAMHNIHSGSGPARREAFKTSR-----NALNVSA 332
Qy 1192 WKQNLPRDEIKLIENICWTLMDRIGY 1217
Db 333 WRHALPPAKIRRVQELCAGALQLLGY 358

RESULT 14
US-08-899-514-2
; Sequence 2, Application US/08899514
; Patent No. 5910581
; GENERAL INFORMATION:
; APPLICANT: HABUCHI, OSAMI
; APPLICANT: FUKUTA, MASAKAZU
; TITLE OF INVENTION: POLYPEPTIDE OF GLYCOSAMINOGLYCAN
; TITLE OF INVENTION: SULFOTRANSFERASE ORIGINATING FROM HUMAN AND DNA CODING
; TITLE OF INVENTION: FOR THE POLYPEPTIDE
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: KNOBBE, MARTENS, OLSON & BEAR, LLP
; STREET: 620 NEWPORT CENTER DRIVE, SIXTEENTH FLOOR
; CITY: NEWPORT BEACH
; STATE: CALIFORNIA
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/899,514
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: DANIEL E ALTMAN
; REGISTRATION NUMBER: 34,115
; REFERENCE/DOCKET NUMBER: TOYAM21.001AUS
```

```
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 714 760 0404
; TELEFAX: 714 760 9502
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 479
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-899-514-2

Query Match      1.8%; Score 120.5; DB 2; Length 479;
Best Local Similarity 16.7%; Pred. No. 0.012;
Matches 71; Conservative 64; Mismatches 152; Indels 137; Gaps 12;

Qy 843 RTEAGSKSLSEGHMDLPDVVITSLPGSABILKQLFFNSSDFLYIRVP-----894
Db 118 RKEEPPRPVAVAGPRRH-----VLLMATRTTSGSFVGEFFNQOQNIIFYLFELPLHIERTV 172
Qy 895 -----TAYIDIPETELEIDSFVDACEWKVSDIRSGH---FRLLRGWLQSLVQ 938
Db 173 SPEPGANNAAGSALVYRDVLKOLFCLDLV--LEHFIPLPEDHLTQFMFRGSSRLCE 230
Qy 939 D-----TKLHLQNIHLHEPNRGKLAQYFAMNKDKKFKKFRRESLPEQRSOMKGAFDRDA 992
Db 231 DPVCTPFVKVPEKYCKNRCGPNLVTLAAEACRRK-----EHMALKAVRIQL 280
Qy 993 EYIRALRHLVYYPARPVLSSSGSWTLKLHFFQEVGLASMRALYIVDRDPRAWIYSMLY 1052
Db 281 EFLQPL-----AEDPRLD-----RVQLVRDPRAVLASRMV 312
Qy 1053 NSKPSLYSLKNVPEHLAKLFKIEGGKCNLNSGYAFEYEPURKLSKSKSNVSLLSHL 1112
Db 313 -----AFAGKYKTWKWLDDGQGLREEVQ 339
Qy 1113 WLANTAALRINTDL-----LPTSQYLVKPEDIVHPPQKTTTIFAFLGIPLSPASLN 1165
Db 340 RLURGNCSIRLSAELGLRQPAWLGRVYMLVRYEDVARGPLQKAREMYPPAGIPTPQVED 399
Qy 1166 QI-----LFATSTNLFYLPYEGEISPTNTNVWKQNLPRDEIKLIENICWTLM 1213
Db 400 WIKNTQAHDGSGIYSTQKN-----SSEQFEKWRFSMPFKLAQVVOAQCPCAMR 449
Qy 1214 RLGY 1217
Db 450 LFGY 453

RESULT 15
US-09-949-016-6813
; Sequence 6813, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6813
; LENGTH: 390
; TYPE: PRT
; ORGANISM: Human
; US-09-949-016-6813
```

```
Query Match      1.8%; Score 119.5; DB 4; Length 390;
Best Local Similarity 21.6%; Pred. No. 0.011;
Matches 86; Conservative 54; Mismatches 163; Indels 95; Gaps 16;

Qy 849 SKKSLSEGHMDLPDVVITSLPGSGAEILKQLFFNSDFLYIRVP-----TA 896
Db 28 SRPGFSPAGGEDRVHVLVLSWSRSGSFLGQLFSQHPDVFYLMPEAWHVWTTLSQGSAA 87
Qy 897 YIDIPETELEIDSFVDACEWKVSDI---RSGHFRLLRGWLQSLVQDTKLHLQNIHLHEP- 952
Db 88 TLHMAVRDLMRISIFL--CDMDVPDAYMPOSRLNLSAFNWTASRA-----LCSP 134
Qy 953 -----NRGKLAQYFAMNKKRKFKRRESLPEQRSMKGAFDRDAEYIRALRRHLVYPS 1007
Db 135 ACSAPPRGTISK-----QDVCKTLCTRQPPFSLAREACRSYSHVVLKEVRFNQLVPL 189
Qy 1008 ARPVLISLSSGSWTLKLHFFQEVLGASMRALYIVDPRAWIYMLYNSKPSLYSLKNVPEH 1067
Db 190 SDPALNL-----RIVHLVRDPRAVLSR--EAAGPILARDNGI--- 225
Qy 1068 LAKLFKIEGGKGCNLSNGYAFYEP---LRKELSKSKSNAVSLSLHMLANTAAALRN 1124
Db 226 -----VLGTNGKW-----VEADPHRLRLIREVCR-----SHVRIAE-AATLK-P 261
Qy 1125 TDLPTSQVLKFDIVHFPQKTTTERIPAFGLIPLSPASLNQILFATSTNLFYLPYEG-E 1183
Db 262 PPFLRGYRLVRFEDLAREPLAEIRALYAFGTGLTLPQLEAWIHNITHGSGIGKPIEAFH 321
Qy 1184 ISPTN-----TNWQNLPRDEIKLIENICWTLMDBRLGY 1217
Db 322 TSSRNARNVSOAMRHLPFTTKILRVQEVCAQALQLGY 359
```

Search completed: June 23, 2005, 08:52:18
Job time : 50.0594 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2005, 08:49:14 ; Search time 154.482 Seconds
(without alignments)

3041.886 Million cell updates/sec

Title: US-10-697-828-15

Perfect score: 6532

Sequence: 1 MPKGGAPPWTMALMFTGHL.....LIENICWTLMRLGYPKFMD 1222

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1717557 seqs, 384547976 residues

Total number of hits satisfying chosen parameters: 1717557

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US10F_PUBCOMB.pep.*
- 19: /cgn2_6/ptodata/1/pubpaa/US11A_PUBCOMB.pep.*
- 20: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
- 21: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 22: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|--------------------|
| 1 | 6532 | 100.0 | 1222 | 16 | US-10-697-828-15 |
| 2 | 6509 | 99.6 | 1222 | 16 | US-10-475-446-4 |
| 3 | 6465 | 99.0 | 1207 | 17 | US-10-479-472A-2 |
| 4 | 5899.5 | 90.3 | 1207 | 16 | US-10-697-828-17 |
| 5 | 5899.5 | 90.3 | 1207 | 17 | US-10-479-472A-4 |
| 6 | 3147 | 48.2 | 596 | 16 | US-10-697-828-9 |
| 7 | 568 | 8.7 | 125 | 11 | US-09-984-429-71 |
| 8 | 568 | 8.7 | 126 | 14 | US-10-150-111-71 |
| 9 | 147 | 2.3 | 411 | 14 | US-10-021-660-128 |
| 10 | 147 | 2.3 | 411 | 15 | US-10-211-462-97 |
| 11 | 147 | 2.3 | 411 | 16 | US-10-408-765A-395 |

| | | | | | |
|----|-------|-----|------|----|----------------------|
| 12 | 147 | 2.3 | 411 | 16 | US-10-723-860-1544 |
| 13 | 145 | 2.2 | 2504 | 9 | US-09-817-514A-8 |
| 14 | 145 | 2.2 | 2504 | 15 | US-10-262-794A-12 |
| 15 | 145 | 2.2 | 2504 | 16 | US-10-754-115-59 |
| 16 | 135 | 2.1 | 1844 | 15 | US-10-262-794A-53 |
| 17 | 134 | 2.1 | 483 | 14 | US-10-212-933-2 |
| 18 | 133 | 2.0 | 484 | 14 | US-10-212-933-4 |
| 19 | 133 | 2.0 | 530 | 16 | US-10-723-860-1409 |
| 20 | 133 | 2.0 | 531 | 9 | US-09-833-790-255 |
| 21 | 133 | 2.0 | 531 | 16 | US-10-755-889-284 |
| 22 | 127 | 1.9 | 1456 | 9 | US-09-870-759-95 |
| 23 | 127 | 1.9 | 1456 | 10 | US-09-751-708A-95 |
| 24 | 127 | 1.9 | 1456 | 15 | US-10-295-027-1217 |
| 25 | 127 | 1.9 | 1456 | 16 | US-10-428-817A-91 |
| 26 | 127 | 1.9 | 1456 | 16 | US-10-723-860-673 |
| 27 | 127 | 1.9 | 1456 | 17 | US-10-937-750A-72 |
| 28 | 126 | 1.9 | 2322 | 17 | US-10-732-923-13997 |
| 29 | 123.5 | 1.9 | 395 | 9 | US-09-927-602-2 |
| 30 | 123.5 | 1.9 | 395 | 15 | US-10-258-080-5 |
| 31 | 123.5 | 1.9 | 395 | 16 | US-10-648-593-159 |
| 32 | 123.5 | 1.9 | 395 | 16 | US-10-697-828-13 |
| 33 | 122.5 | 1.9 | 479 | 13 | US-10-087-192-126 |
| 34 | 120 | 1.8 | 2360 | 17 | US-10-732-923-13960 |
| 35 | 119.5 | 1.8 | 390 | 9 | US-09-927-602-4 |
| 36 | 119.5 | 1.8 | 390 | 16 | US-10-697-828-8 |
| 37 | 119 | 1.8 | 481 | 13 | US-10-087-192-123 |
| 38 | 118.5 | 1.8 | 418 | 9 | US-09-927-602-5 |
| 39 | 118 | 1.8 | 395 | 16 | US-10-697-828-7 |
| 40 | 118 | 1.8 | 395 | 16 | US-10-841-707-8 |
| 41 | 115.5 | 1.8 | 386 | 14 | US-10-427-631-11 |
| 42 | 115 | 1.8 | 893 | 13 | US-10-014-436-4 |
| 43 | 115 | 1.8 | 1938 | 13 | US-10-014-436-2 |
| 44 | 113.5 | 1.7 | 1497 | 16 | US-10-437-963-138435 |
| 45 | 113 | 1.7 | 395 | 9 | US-09-927-602-3 |

ALIGNMENTS

RESULT 1
US-10-697-828-15
; Sequence 15, Application US/10697828
; Publication No. US20040185546A1
; GENERAL INFORMATION:
; APPLICANT: Rosen, Steven
; APPLICANT: Lee, Jin Kyu
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: Novel Glycosyl Sulfotransferases GST-4alpha, GST-4beta, & GST-6
; FILE REFERENCE: UCAL-138DIV
; CURRENT FILING DATE: 2003-10-29
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: 09/593,828
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: 60/144,694
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 1222
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-697-828-15

| | | | | | | | |
|-----------------------|--------|---|------|------------|----|--------|------|
| Query Match | 100.0% | Score | 6532 | DB | 16 | Length | 1222 |
| Best Local Similarity | 100.0% | Pred. No. | 0 | | | | |
| Matches | 1222 | Conservative | 0 | Mismatches | 0 | Indels | 0 |
| Gaps | 0 | | | | | | |
| Qy | 1 | MPKGGAPPWTMALMFTGHLFLALLMFAFSTFEESVSNYSWAVFTDDIDQFKTKQVDF | 60 | | | | |
| Db | 1 | MPKGGAPPWTMALMFTGHLFLALLMFAFSTFEESVSNYSWAVFTDDIDQFKTKQVDF | 60 | | | | |
| Qy | 61 | RPNOKLKKSMLHPSLYFDAGBIQAKRSASHLHLFRATISAVTVMLSNPTYYLPPPKH | 120 | | | | |

Db 61 RPNQKKSMLHPSLYFDAGEIQAMRQKSRASHLHLFRAIRSAVTVMLSNPTYYLPPPKH 120
Qy 121 ADFAAKWNEIYGNLPLALYCLLPCKDKVAFVLEIYMDRMVGYKDWLVENAPGDVPI 180
Db 121 ADFAAKWNEIYGNLPLALYCLLPCKDKVAFVLEIYMDRMVGYKDWLVENAPGDVPI 180
Qy 181 GHSUTGFATADFLYLNLLDNRHROKYLEKIWIITEEMVEYSKVRSGKQLLHNHOATNMI 240
Db 181 GHSUTGFATADFLYLNLLDNRHROKYLEKIWIITEEMVEYSKVRSGKQLLHNHOATNMI 240
Qy 241 ALLTGALVTGVDGKSKANIWKQAVVDMKTMFLNHI VDGSLDEGVAYGSYTAKSQY 300
Db 241 ALLTGALVTGVDGKSKANIWKQAVVDMKTMFLNHI VDGSLDEGVAYGSYTAKSQY 300
Qy 301 VFLAQRHFNINLNNLWKHFWFYATLLPGFORTVGIADSNYNWFYGPESQLVFLDKF 360
Db 301 VFLAQRHFNINLNNLWKHFWFYATLLPGFORTVGIADSNYNWFYGPESQLVFLDKF 360
Qy 361 ILKNGAGNLAQAQIRKRPKDGMPVSTAQRWSTLHTEYIWDYDQLTPOPPADYGTAKIH 420
Db 361 ILKNGAGNLAQAQIRKRPKDGMPVSTAQRWSTLHTEYIWDYDQLTPOPPADYGTAKIH 420
Qy 421 TFPNMGVVITYGAGLPNTQTNTFVSFKSGKLGRAVYDVIHFQPYSWIDGWRSPNPGHEHP 480
Db 421 TFPNMGVVITYGAGLPNTQTNTFVSFKSGKLGRAVYDVIHFQPYSWIDGWRSPNPGHEHP 480
Qy 481 DONSFTFAPNGQVPSBALYGPKLSHLNNVLVFPAPSSQCNKPWEGQLGECQWLKWTG 540
Db 481 DONSFTFAPNGQVPSBALYGPKLSHLNNVLVFPAPSSQCNKPWEGQLGECQWLKWTG 540
Qy 541 EEUVDAAAGEIITASOHGEMFVSGEAYSASSAMRLKSVYRALLLNSQTLVVVDHIERQ 600
Db 541 EEUVDAAAGEIITASOHGEMFVSGEAYSASSAMRLKSVYRALLLNSQTLVVVDHIERQ 600
Qy 601 EDSPIVNSVFAFFHNLDIDFKYIPYKFNRYNGAMMDVDAHYKMFVDDHGNSPMASIQE 660
Db 601 EDSPIVNSVFAFFHNLDIDFKYIPYKFNRYNGAMMDVDAHYKMFVDDHGNSPMASIQE 660
Qy 661 AEQAAEFKRWQTVNVTWFQWESITIRIAYFYGPYINVSCTRIDSNFGLQISLNVNN 720
Db 661 AEQAAEFKRWQTVNVTWFQWESITIRIAYFYGPYINVSCTRIDSNFGLQISLNVNN 720
Qy 721 TEHVVSIVTDVHNLKTRFNLYLGGFGFASVADQOITRFGLTQAIKVPVRHDIRIIPFPG 780
Db 721 TEHVVSIVTDVHNLKTRFNLYLGGFGFASVADQOITRFGLTQAIKVPVRHDIRIIPFPG 780
Qy 781 KFNIAVGLILCISLVILTQWRFYLSFRKLMRWILILVIALWFIELLDVNSTCSQPICAK 840
Db 781 KFNIAVGLILCISLVILTQWRFYLSFRKLMRWILILVIALWFIELLDVNSTCSQPICAK 840
Qy 841 WTRTEAGSKKSLSEGGHMDLPDVVITSLPGSGAELKQLFNSSDFLYIRVPTAYIDI 900
Db 841 WTRTEAGSKKSLSEGGHMDLPDVVITSLPGSGAELKQLFNSSDFLYIRVPTAYIDI 900
Qy 901 PETELEIDSFVDACEWKVSDIRSGHFRLLRGWLSQVQDTKLHLQNLHLHHPNKGKLAQY 960
Db 901 PETELEIDSFVDACEWKVSDIRSGHFRLLRGWLSQVQDTKLHLQNLHLHHPNKGKLAQY 960
Qy 961 FANMKDKRFRKRESLPEQSRQMKGFADRAEYIRALRRHLVYYPGARPVLSLSSGSWT 1020
Db 961 FANMKDKRFRKRESLPEQSRQMKGFADRAEYIRALRRHLVYYPGARPVLSLSSGSWT 1020
Qy 1021 LKHLFFOEVLGASMRALYIVRDPRAWIYSMLYNSKPSLYSLKNVPEHLAKLFIEGGK 1080
Db 1021 LKHLFFOEVLGASMRALYIVRDPRAWIYSMLYNSKPSLYSLKNVPEHLAKLFIEGGK 1080
Qy 1081 CNLNSGVAFEYELRKEISKSNVSLSHLWLANTAARLNTDLLPTSQYLKVFEDI 1140
Db 1081 CNLNSGVAFEYELRKEISKSNVSLSHLWLANTAARLNTDLLPTSQYLKVFEDI 1140
Qy 1141 VHPFQKTERIFAFGLIGPLSPASLNQILFATSTNLFLPYEGEISPTNTNWKQNLPRDE 1200
Db 1141 VHPFQKTERIFAFGLIGPLSPASLNQILFATSTNLFLPYEGEISPTNTNWKQNLPRDE 1200

Qy 1201 IKLIENICWTMLDRLCYBPKFMD 1222
Db 1201 IKLIENICWTMLDRLCYBPKFMD 1222
RESULT 2
US-10-475-446-4
; Sequence 4, Application US/10475446
; Publication No. US20040198651A1
; GENERAL INFORMATION:
; APPLICANT: KLAMMER, Aaron A.; HAFALIA, April J.A.
; APPLICANT: DUGGAN, Brendon M.; WARREN, Bridget A.
; APPLICANT: EMERLING, Brooke M.; TRIBULEY, Catherine M.
; APPLICANT: ARVIZU, Chandra S.; HONCHELL, Cynthia D.
; APPLICANT: NGUYEN, Dannel B.; KALLICK, Deborah A.
; APPLICANT: YUE, Henry; AU-YOUNG, Janice K.
; APPLICANT: RAMKUMAR, Jayalaxmi; LI, Joana X.
; APPLICANT: THANGAVELU, Kavitha; GIETZEN, Kimberly J.
; APPLICANT: DING, Li; BAUGHN, Mariah R.
; APPLICANT: YAO, Monique G.; CHAWLA, Narinder K.
; APPLICANT: MASON, Patricia M.; LAL, Preeti G.
; APPLICANT: GRAU, Richard C.; REDDY, Roopa M.
; APPLICANT: BECHA, Shanya D.; KAREHT, Stephanie K.
; APPLICANT: RICHARDSON, Thomas W.; TRAN, Uyen K.
; APPLICANT: ELLIOTT, Vicki S.; TANG, Y. Tom
; APPLICANT: AZIMZAI, Yalda; LU, Yan
; APPLICANT: XU, Yuming
; TITLE OF INVENTION: SECRETED PROTEINS
; FILE REFERENCE: PF-0949 USN
; CURRENT APPLICATION NUMBER: US/10/475,446
; CURRENT FILING DATE: 2003-10-20
; PRIOR APPLICATION NUMBER: PCT/US02/12464
; PRIOR FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US 60/285,207
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 60/287,114
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/288,640
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: US 60/290,516
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/292,184
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 60/343,553
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: US 60/357,002
; PRIOR FILING DATE: 2002-02-13
; PRIOR APPLICATION NUMBER: US 60/358,279
; PRIOR FILING DATE: 2002-02-20
; PRIOR APPLICATION NUMBER: US 60/366,041
; PRIOR FILING DATE: 2002-03-19
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PERL Program
; SEQ ID NO 4
; LENGTH: 1222
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 3441255CD1
US-10-475-446-4

Query Match 99.6%; Score 6509; DB 16; Length 1222;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1219; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 MPKGGAPPITAMLFTHLLFLALLMPAFSTFEESVSNYSEWAVFTDDIDQFKTKQVDF 60
Db 1 MPKGGAPPITAMLFTHLLFLALLMPAFSTFEESVSNYSEWAVFTDDIDQFKTKQVDF 60
Qy 61 RPNQKKSMLHPSLYFDAGEIQAMRQKSRASHLHLFRAIRSAVTVMLSNPTYYLPPPKH 120

```
Db 61  RPNQKKSLHPSLYFDAGEIQAMROKSRASHLHLFRAIRSATVVMLSNPTYYLPPPKH 120
Qy 121  ADFAAKNEIYGNLPLALYCLLCPEDKVAFEVLEVMRVMGVYKDWLVENAPGDEVPI 180
Db 121  ADFAAKNEIYGNLPLALYCLLCPEDKVAFEVLEVMRVMGVYKDWLVENAPGDEVPI 180
Qy 181  GHSLTGPATAPDFLYNLDDNHRQKYLEKIWIITEEMYESKVRSGKQLLHNHQAENMI 240
Db 181  VHSLTGPATAPDFLYNLDDNHRQKYLEKIWIITEEMYESKVRSGKQLLHNHQAENMI 240
Qy 241  ALLTGALVTGVDGSKKANIWKQAVVDMVMEKTMFLLNHI VDGSLDEGVAYGSYTAQSVTQY 300
Db 241  ALLTGALVTGVDGSKKANIWKQAVVDMVMEKTMFLLNHI VDGSLDEGVAYGSYTAQSVTQY 300
Qy 301  VFLAQRHFNINLNNLWKHFWFYATLLPGFQRTVGIADSNYNWFYGPESQLVFLDKF 360
Db 301  VFLAQRHFNINLNNLWKHFWFYATLLPGFQRTVGIADSNYNWFYGPESQLVFLDKF 360
Qy 361  ILKNGAGNWLAAQIIRKHPKDGMPVSTAQRMSTLHTEYIWDYDQLTPQPPADYGTAKIH 420
Db 361  ILKNGAGNWLAAQIIRKHPKDGMPVSTAQRMSTLHTEYIWDYDQLTPQPPADYGTAKIH 420
Qy 421  TFPNMGVVTYGAGLPNTQTNTFVSFKSGKLGGRVAVYDI VHFQPSYDWDGWSRFPNGHEHP 480
Db 421  TFPNMGVVTYGAGLPNTQTNTFVSFKSGKLGGRVAVYDI VHFQPSYDWDGWSRFPNGHEHP 480
Qy 481  DONSFTFAPNGQVFSBALYGPKLSHLNNVLVPAPSSQCNKPWGQLGCECAQWLKWTG 540
Db 481  DONSFTFAPNGQVFSBALYGPKLSHLNNVLVPAPSSQCNKPWGQLGCECAQWLKWTG 540
Qy 541  EEVGDAAGEIITASQHGEMVFSGEAVSAYSSAMRLKSVYRALLLLNSQTLVVVDHIERQ 600
Db 541  EEVGDAAGEIITASQHGEMVFSGEAVSAYSSAMRLKSVYRALLLLNSQTLVVVDHIERQ 600
Qy 601  EDSINSVSAFFHNLDIDFKYIPYKFNRYNGAMVDVDAHYKWFVDDHNGNSPMASIQE 660
Db 601  EDSINSVSAFFHNLDIDFKYIPYKFNRYNGAMVDVDAHYKWFVDDHNGNSPMASIQE 660
Qy 661  AEQAAEFKKWTOFVNVTQOMESTITRIAYVYGPYINVSCRFIDSSNPGLOISLVN 720
Db 661  AEQAAEFKKWTOFVNVTQOMESTITRIAYVYGPYINVSCRFIDSSNPGLOISLVN 720
Qy 721  TEHVVSIVTDYHNLKTRFNLYGCGFASVADQGI TRFGLGTQAI VPKVRHDIRIIFPGF 780
Db 721  TEHVVSIVTDYHNLKTRFNLYGCGFASVADQGI TRFGLGTQAI VPKVRHDIRIIFPGF 780
Qy 781  KFNIAVGLILCISVLITFQWRFYLSFRKLMRWILILVIALWFIELDVWSTCSQPICAK 840
Db 781  KFNIAVGLILCISVLITFQWRFYLSFRKLMRWILILVIALWFIELDVWSTCSQPICAK 840
Qy 841  WTRTEAGSKSLSEGHMDLPDVVITSLPGSABILKQLFNNSDPLYIRVPTAYIDI 900
Db 841  WTRTEAGSKSLSEGHMDLPDVVITSLPGSABILKQLFNNSDPLYIRVPTAYIDI 900
Qy 901  PETELEIDSFVDACEWKVSDIRSGHFRLLRGLQSLVQDTKLHQN IHLHPNKGKLAQY 960
Db 901  PETELEIDSFVDACEWKVSDIRSGHFRLLRGLQSLVQDTKLHQN IHLHPNKGKLAQY 960
Qy 961  FAMNKKRKPRESLPEQORSQMGAPDRDAEYIRALRRHLVYVPSARPVLSLSSGWT 1020
Db 961  FAMNKKRKPRESLPEQORSQMGAPDRDAEYIRALRRHLVYVPSARPVLSLSSGWT 1020
Qy 1021  LKLFHQEVLGASMRALYIVDPDRAWIYSMLYNKPSLYSLKNVPEHLAKLFKEGGKGK 1080
Db 1021  LKLFHQEVLGASMRALYIVDPDRAWIYSMLYNKPSLYSLKNVPEHLAKLFKEGGKGK 1080
Qy 1081  CNLNSGYAFEPYELRKLKLSKSNVSLLSHLWLANATAALRINTDLLPTSQYLKVPEDI 1140
Db 1081  CNLNSGYAFEPYELRKLKLSKSNVSLLSHLWLANATAALRINTDLLPTSQYLKVPEDI 1140
Qy 1141  VHFQKTTTERIFAFILGIPLSPASLNQILFATSTNLFYLPYGEISPTNTNWKQNLPRDE 1200
Db 1141  VHFQKTTTERIFAFILGIPLSPASLNQILFATSTNLFYLPYGEISPTNTNWKQNLPRDE 1200
```

```
Qy 1201 IKLIENICWTMLMDRLGYPKPM 1222
Db 1201 IKLIENICWTMLMDRLGYPKPM 1222
```

RESULT 3

```
US-10-479-472A-2
; Sequence 2, Application US/10479472A
; Publication No. US20050118581A1
; GENERAL INFORMATION:
; APPLICANT: DEL-FAVERO, JURGEN PETER LODE
; APPLICANT: VAN BROECKHOVEN, CHRISTINE
; TITLE OF INVENTION: NOVEL BRAIN EXPRESSED GENE AND PROTEIN ASSOCIATED WITH
; TITLE OF INVENTION: BIPOLAR DISORDER
; FILE REFERENCE: JAB-1711
; CURRENT APPLICATION NUMBER: US/10/479,472A
; CURRENT FILING DATE: 2003-12-01
; PRIOR APPLICATION NUMBER: PCT/EP02/06316
; PRIOR FILING DATE: 2002-06-06
; PRIOR APPLICATION NUMBER: EP 01202214.1
; PRIOR FILING DATE: 2001-06-11
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patencin ver. 3.2
; SEQ ID NO 2
; LENGTH: 1212
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-479-472A-2
```

```
Query Match 99.0%; Score 6465; DB 17; Length 1212;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1211; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy 11  MALMFTGHLFLALLMPAFSTFESVSNYSEWAVFTDDIDQFKTKQVDPNPKNLKKSM 70
Db 1  MALMFTGHLFLALLMPAFSTFESVSNYSEWAVFTDDIDQFKTKQVDPNPKNLKKSM 60
Qy 71  LHPSLYFDAGEIQAMROKSRASHLHLFRAIRSATVVMLSNPTYYLPPPKHADFAAKWNEI 130
Db 61  LHPSLYFDAGEIQAMROKSRASHLHLFRAIRSATVVMLSNPTYYLPPPKHADFAAKWNEI 120
Qy 131  YGNLPLALYCLLCPEDKVAFEVLEVMRVMGVYKDWLVENAPGDEVPIGHSITGFATA 190
Db 121  YGNLPLALYCLLCPEDKVAFEVLEVMRVMGVYKDWLVENAPGDEVPIGHSITGFATA 180
Qy 191  FDFLYNLDDNHRQKYLEKIWIITEEMYESKVRSGKQLLHNHQAENMIALLTGALVTG 250
Db 181  FDFLYNLDDNHRQKYLEKIWIITEEMYESKVRSGKQLLHNHQAENMIALLTGALVTG 240
Qy 251  VDKGSKANIWKQAVVDMVMEKTMFLLNHI VDGSLDEGVAYGSYTAQSVTQYVFLAQRHFN 310
Db 241  VDKGSKANIWKQAVVDMVMEKTMFLLNHI VDGSLDEGVAYGSYTAQSVTQYVFLAQRHFN 300
Qy 311  NNLDNNWLKWHFWFYATLLPGFQRTVGIADSNYNWFYGPESQLVFLDKFLLKNGAGNWL 370
Db 301  NNLDNNWLKWHFWFYATLLPGFQRTVGIADSNYNWFYGPESQLVFLDKFLLKNGAGNWL 360
Qy 371  AQQIRKHPKDGMPVSTAQRMSTLHTEYIWDYDQLTPQPPADYGTAKIHFPNMGVVTY 430
Db 361  AQQIRKHPKDGMPVSTAQRMSTLHTEYIWDYDQLTPQPPADYGTAKIHFPNMGVVTY 420
Qy 431  GAGLPNTQTNTFVSFKSGKLGGRVAVYDI VHFQPSYDWDGWSRFPNGHEHPDQNSFTFAPN 490
Db 421  GAGLPNTQTNTFVSFKSGKLGGRVAVYDI VHFQPSYDWDGWSRFPNGHEHPDQNSFTFAPN 480
Qy 491  GQVFSVSEALYGPKLSHLNNVLVPAPSSQCNKPWGQLGCECAQWLKWTGEEVGDAAAGEI 550
Db 481  GQVFSVSEALYGPKLSHLNNVLVPAPSSQCNKPWGQLGCECAQWLKWTGEEVGDAAAGEI 540
Qy 551  ITASQHGEMVFSGEAVSAYSSAMRLKSVYRALLLLNSQTLVVVDHIERQSDPSINSVSA 610
Db 541  ITASQHGEMVFSGEAVSAYSSAMRLKSVYRALLLLNSQTLVVVDHIERQSDPSINSVSA 600
```

611 FFHNLIDDFKYPYKFMNRYNGAMMDVMDAHYKMFDFDHGNSPMASIOEAEQAAEFKKR 670
601 FFHNLIDDFKYPYKFMNRYNGAMMDVMDAHYKMFDFDHGNSPMASIOEAEQAAEFKKR 660
671 WTQFVNVTQFMESTTIRIAYVFYGPYINVSCEFIDSSNFGLOISLVNNTTEHVVSIVTD 730
661 WTQFVNVTQFMEPTITRIAYVFYGPYINVSCEFIDSSNFGLOISLVNNTTEHVVSIVTD 720
731 YHNLKTRFNVLGGGFPASVADQGOITRFGLTQAIIVKPVHRDRIIPPGFKFNIAVGLIL 790
721 YHNLKTRFNVLGGGFPASVADQGOITRFGLTQAIIVKPVHRDRIIPPGFKFNIAVGLIL 780
791 CISLVILTFQWRFLVSRKLMRWILILVIALWFIELLDVWSTCSQPICAKWTTEAGSK 850
781 CISLVILTFQWRFLVSRKLMRWILILVIALWFIELLDVWSTCSQPICAKWTTEAGSK 840
851 KSLSEGHMDLPDVVITSLPGSGAEILKQLFNSSDFLYIRVPTAYIDIPETELEDISF 910
841 KSLSEGHMDLPDVVITSLPGSGAEILKQLFNSSDFLYIRVPTAYIDIPETELEDISF 900
911 VDACEWKVSDIRSGHFRLLRGWLQSLVQDTKHLQNIHLHEPNRGKLAQYFAMNKDKKKR 970
901 VDACEWKVSDIRSGHFRLLRGWLQSLVQDTKHLQNIHLHEPNRGKLAQYFAMNKDKKKR 960
971 FKRESLPEORSQMGAFDRDAEYIRALRRHLVYPSARPVLSLSSGWTCLKHFFQEVIL 1030
961 FKRESLPEORSQMGAFDRDAEYIRALRRHLVYPSARPVLSLSSGWTCLKHFFQEVIL 1020
1031 GASMRALYIVRDPRAWIYMLYNSKPSLSYSLKNVPBHLAKLFIIEGKGKCNLSGYAFE 1090
1021 GASMRALYIVRDPRAWIYMLYNSKPSLSYSLKNVPBHLAKLFIIEGKGKCNLSGYAFE 1080
1091 YEPLRKELSKSKNAVSLLSHLMLANTAAALRINTDILLPSYQLVKPEDIIVHPQKTTT 1150
1081 YEPLRKELSKSKNAVSLLSHLMLANTAAALRINTDILLPSYQLVKPEDIIVHPQKTTT 1140
1151 IFAPLGIPLSPASINQILFATSNLFLYPVEGEISPTNTNVKONLPROEIKLIENICWT 1210
1141 IFAPLGIPLSPASINQILFATSNLFLYPVEGEISPTNTNVKONLPROEIKLIENICWT 1200
1211 LMDRLGYPKFMD 1222
1201 LMDRLGYPKFMD 1212

RESULT 4

US-10-697-828-17
; Sequence 17, Application US/10697828
; Publication No. US20040185546A1
; GENERAL INFORMATION:
; APPLICANT: Rosen, Steven
; APPLICANT: Lee, Jin Kyu
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: Novel Glycosyl Sulfotransferases GST-4alpha, GST-4beta, & GST-6
; FILE REFERENCE: UCAL-138DIV
; CURRENT APPLICATION NUMBER: US/10/697,828
; PRIOR FILING DATE: 2003-10-29
; PRIOR APPLICATION NUMBER: 09/593,828
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: 60/144,694
; PRIOR FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 1207
; TYPE: PRT
; ORGANISM: mouse
US-10-697-828-17

Query Match

Best Local Similarity 90.3%; Score 5899.5; DB 16; Length 1207;

Matches 1084; Conservative 68; Mismatches 55; Indels 5; Gaps 1;

11 MALMFTGHLFLALLMEAFSTPEESVSNYSEWAVFTDDIDQFKTKOKVQDFRPNOKLKSM 70
1 MAFMTEHLLFLUTLMWCSFSTCEESVSNYSEWAVFTDDIQWLKSKQIQDFKUNRR----- 55
71 LHPSLYFDAGEIQAMRQKSRASHLHLFRAIRSAVTVMLSNPTYIPLPPKHADFAAKWNEI 130
56 LHPNLYFDAGDIQTLKQKSRSTSLHAI FRAIKSAVTIMLSNPSYIPLPPKHAFSAKWNNEI 115
131 YGNLPLPPLALYCLLPEDKVAPEFVLEWDMRWGVKDMVLVENAPGDEVPIGHSLTGFATA 190
116 YGNLPLPPLALYCLLPEDKVAPEFVMEYMDRMVSYKDMVLVENAPGDEVPGVHSLTGFATA 175
191 FDFLNLNLDNHRROKYLEKIWIITEEMEYKSVRSWKGKOLLHNHOATNMIALLTGALVTG 250
176 FDFLNLNLDNHRROKYLEKIWIITEEMEYKSVRSWKGKOLLHNHOATNMIALLTGALVTG 235
251 VDKGSKANIWKQAVVDVMEKTMFLNLNHIVDGSLDEGVAYGYTAKSVTYQYVFLAQRHFI 310
236 VDKGSKANIWKQAVVDVMEKTMFLNLNHIVDGSLDEGVAYGYTAKSVTYQYVFLAQRHFI 295
311 NNLDNNWLKQHPFWFYATLLPGQRTVGIADSNYNWFGPESQLVFLDKFILQNGAGNWL 370
296 NNFDNNWLKQHPFWFYATLLPGQRTVGIADSNYNWFGPESQLVFLDKFILQNGAGNWL 355
371 AQOIRKRRPKQGPMPVSTAQRWSTLHTEYIWDYDQDLPQPPADYGTAKIHTFPNKGVTY 430
356 AQOIRKRRPKQGPMPVSTAQRWSTLHTEYIWDYDQDLPQPPADYGTAKIHTFPNKGVTY 415
431 GAGLNTQNTNTFVSPKSGKLGGRAVYDIVHFQYPSWIDGWSFNGPHEHPDQNSTFPAPN 490
416 GAGLNTQNTNTFVSPKSGKLGGRAVYDIVHFQYPSWIDGWSFNGPHEHPDQNSTFPAPN 475
491 GQVFVSEALYGPKLHNLNVLVPAPSSQCNKPWEGQIGECAQWLKWTGEVGAAGEI 550
476 GQVFVSEALYGPKLHNLNVLVPAPSSQCNKPWEGQIGECAQWLKWTGEVGAAGEV 535
551 ITASQHGEMVFGSGAVSAYSSAMELKSVYRALLLNSQTLVVDHIERQDPSINSVA 610
536 ITAAQHGORMFVSGGAVSAYSSAMELKSVYRALLLNSQTLVVDHIERQDPSINSVA 595
611 FPHNLIDDFKYPYKFMNRYNGAMMDVMDAHYKMFDFDHGNSPMASIOEAEQAAEFKKR 670
596 FPHNLIDDFKYPYKFMNRYNGAMMDVMDAHYKMFDFDHGNSPMASIOEAEQAAEFKKR 655
671 WTQFVNVTQFMESTTIRIAYVFYGPYINVSCEFIDSSNFGLOISLVNNTTEHVVSIVTD 730
656 WTQFVNVTQFMESTTIRIAYVFYGPYINVSCEFIDSSNFGLOISLVNNTTEHVVSIVTD 715
731 YHNLKTRFNVLGGGFPASVADQGOITRFGLTQAIIVKPVHRDRIIPPGFKFNIAVGLIL 790
716 YONLKSRYSLGFGGFPASVADQGOITRFGLTQAIIVKPVHRDRIIPPGFKFNIAVGLIL 775
791 CISLVILTFQWRFLVSRKLMRWILILVIALWFIELLDVWSTCSQPICAKWTTEAGSK 850
776 CISLVILTFQWRFLVSRKLMRWILILVIALWFIELLDVWSTCSQPICAKWTTEAGSK 835
851 KSLSEGHMDLPDVVITSLPGSGAEILKQLFNSSDFLYIRVPTAYIDIPETELEDISF 910
836 KYMISEGHHVDLPNVIIITSLPGSGAEILKQLFNSSDFLYIRVPTAYIDIPETELEDISF 895
911 VDACEWKVSDIRSGHFRLLRGWLQSLVQDTKHLQNIHLHEPNRGKLAQYFAMNKDKKKR 970
896 VDACEWKVSDIRSGHFRLLRGWLQSLVQDTKHLQNIHLHEPNRGKLAQYFAMNKDKKKR 955
971 FKRESLPEORSQMGAFDRDAEYIRALRRHLVYPSARPVLSLSSGWTCLKHFFQEVIL 1030
956 LKRESLQDQRSRIKGPFDRAEYIRALRRHLVYPSARPVLSLSSGWTCLKHFFQEVIL 1015
1031 GASMRALYIVRDPRAWIYMLYNSKPSLSYSLKNVPBHLAKLFIIEGKGKCNLSGYAFE 1090
1016 GTSMRALYIVRDPRAWIYMLYNSKPSLSYSLKNVPBHLAKLFIIEGKGKCNLSGYAFE 1075

```
Qy 1091 YELRLKELSKSNVALLSHLWANTAAALRINTDLPSTSYQLVKPEDIVHFPQKTER 1150
Db 1076 YESLKELEISQSNALSLSHLWANTAAALRINTDLPSTNYHLVKPEDIVHFPQKTER 1135
Qy 1151 IFAPLGIPLSPASINQILPATSTNLFYLPYEGEISPTNTNWKONLPRDEIKLIENICWT 1210
Db 1136 IFAPLGIPLSPASINQILPATSTNLFYLPYEGEISPTNTNWKONLPRDEIKLIENICWT 1195
Qy 1211 LMDRLGYPKFMD 1222
Db 1196 LMDHLGYPKFMD 1207

RESULT 5
US-10-479-472A-4
; Sequence 4, Application US/10479472A
; Publication No. US20050118581A1
; GENERAL INFORMATION:
; APPLICANT: DEL-FAVERO, JURGEN PETER LODE
; APPLICANT: VAN BROECKHOVEN, CHRISTINE
; TITLE OF INVENTION: NOVEL BRAIN EXPRESSED GENE AND PROTEIN ASSOCIATED WITH
; TITLE OF INVENTION: BIPOLAR DISORDER
; FILE REFERENCE: JAB-1711
; CURRENT APPLICATION NUMBER: US/10/479,472A
; CURRENT FILING DATE: 2003-12-01
; PRIOR APPLICATION NUMBER: PCT/EP02/06316
; PRIOR FILING DATE: 2002-06-06
; PRIOR APPLICATION NUMBER: EP 01202214.1
; PRIOR FILING DATE: 2001-06-11
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 4
; LENGTH: 1207
; TYPE: PRT
; ORGANISM: Mus sp.
US-10-479-472A-4

Query Match 90.3%; Score 5899.5; DB 17; Length 1207;
Best Local Similarity 89.4%; Pred. No. 0;
Matches 1084; Conservative 68; Mismatches 55; Indels 5; Gaps 1;

Qy 11 MALMFTGHLFLALLMFAFSTFEESVSNYSWAYFTDDIDQFKTKQVQDPRPNQKLKSM 70
Db 1 MAPMTEHLLFLTLMMCSFSTCEESVSNYSWAYFTDDIQWLKSKQIQDPKLNRR----- 55
Qy 71 LHPSLYFDAGEIQAMROKRSASHLHLFRAIRSAVTMLSNPTTYLPPPKHADPAKWEI 130
Db 56 LHPNLYFDAGDIQTLKQKSRSTSHLHIFRAIKSAVTIMLSNPSYLLPPPKHAEFAAKWEI 115
Qy 131 YGNLPLALYCLLCPEDKVAFAFVLEYMDRMVGYKDWLVENAPGDEVPICHSITGPATA 190
Db 116 YGNLPLALYCLLCPEDKVAFAFVLEYMDRMVGYKDWLVENAPGDEVPVGSHTGPATA 175
Qy 191 FDFLYNLDDHRRQKYLEIKIWIITEEMVEYSKVRSGKQLLHNRHQATNMIALLTGALVTG 250
Db 176 FDFLYNLGNQKQKYLEIKIWIITEEMVEYSKIRSGKQLLHNRHQATNMIALLTGALVTG 235
Qy 251 VDKGSKANIWKQAVDVMEKTMFLNLHIVDGSGLDEGVAYGSAKSTQYVFLAQRHFI 310
Db 236 VDKGSKANIWKQAVDVMEKTMFLNLHIVDGSGLDEGVAYGSAKSTQYVFLAQRHFI 295
Qy 311 NNLONNLKQHFYFYATLLPGFQRTVGIADSNYNWFGPESQVFLDKFTLKNAGNWL 370
Db 296 NFNONNLKQHFYFYATLLPGYQRTVGIADSNYNWFGPESQVFLDKFTLQNGAGNWL 355
Qy 371 AQQIRKRPKDGMPVPSTQARWSLHTEYIWIYDLPQTPQPPADYCTAKIHTFPNKGVVY 430
Db 356 AQQIRKRPKDGMPVPSTQARWSLHTEYIWIYDPLTPQPPVDPFGTAKWHTFPNKGVVY 415
Qy 431 GAGLPNTQNTFVSFKSGKLGGRVYDVIHVPQYSWIDGWSFNPGEHPHQNSFTFAPN 490
Db 416 GGGLPNTQNTFVSFKSGKLGGRVYDVIHVPQYSWIDGWSFNPGEHPHQNSFTFAPN 475
```

```
Qy 491 GOVTVSEALYGPKLSHLNNLVLPAPSPSSOCNKPWEGQLGECQWLKWTGEEVGDAAGEI 550
Db 476 GOVTVSEALYGPKLSHLNNLVLPAPSPSSOCNKPWEGQLGECQWLKWTGEEVGDAAAGEV 535
Qy 551 ITASQHGEMVFGSEAVSAVSSAMRLKSVYRALLLNQTLVVVDHIERQEDSPINSVSA 610
Db 536 ITAAQHGDRMFVFGSEAVSAVSSAMRLKSVYRALLLNQTLVVVDHIERQETSPINSVSA 595
Qy 611 FFHNLDDIDFKYIPYKFMNRYNGAMDMVDAYHKMFDFDHHGNSPMASIQBAEQAAEFKKR 670
Db 596 FFHNLDDIDFKYIPYKFMNRYNGAMDMVDAYHKMFDFDHHGNSPMVANIQBAEQAAEFKKR 655
Qy 671 WTOPVNVTFOMESTITRIAYFYGPYINVSRCRPIIDSSNPGLOISLNNVTEHVSVITD 730
Db 656 WTOPVNVTFHMESTITRIAYFYGPYINVSRCRPIIDSSNPGLOISLNNVTEHVSIVT 715
Qy 731 YHNLKTRFNYLFGFGFASVADQGITRFGIGTQAIKVPVRHRIIPFPKFNIAVGLIL 790
Db 716 YQNLKGRFSYLGFGFGFASVANQGITRFGIGTQAIKVPVRHRIIPFPKFNIAVGLIL 775
Qy 791 CISLVILTFQWRFYLSFRKLMRWILILVIALMFIELLDVMTSCSQPICAKWTRTEAESGK 850
Db 776 CISLVILTFQWRFYLSFRKLMRCVLILVIALMFIELLDVMTSCSQPICAKWTRTEAKANE 835
Qy 851 KSLSSGHHMDLPDVVITSLPGSGAEILKQLPFNSDFLIRVPTAVIDIPETELEIDSF 910
Db 836 KVMISGHHVDLPNVITSLPGSGAEILKQLPFNSDFLIRVPTAVIDIPETELEIDSF 895
Qy 911 VDACEWKVSDIRSGHFLRGLQSLVQDTKLHLQNLHLHETSKSLAQYFTTNKDKKRR 970
Db 896 VDACEWKVSDIRSGHFLRGLQSLVQDTKLHLQNLHLHETSKSLAQYFTTNKDKKRR 955
Qy 971 FKRESLPEQORSQMGAFDRDAEYIRALRRHLVYPSARPVLSLSSGSWTLKLHFFQEV 1030
Db 956 LKRESLQDQSRIRKGFDRDAEYIRALRRHLVYPSARPVLSLSSGSWTLKLHFFQEV 1015
Qy 1031 GASMRALYIVRDPRAWIYSLYKSPSLYSLKNVPEHLAKFKITEGKGKCNLSGVAFE 1090
Db 1016 GTSMRALYIVRDPRAWIYSLYKSPSLYSLKNVPEHLAKFKITEGKGKCNLSGVAFE 1075
Qy 1091 YELRLKELSKSNVALLSHLWANTAAALRINTDLPSTSYQLVKPEDIVHFPQKTER 1150
Db 1076 YESLKELEISQSNALSLSHLWANTAAALRINTDLPSTNYHLVKPEDIVHFPQKTER 1135
Qy 1151 IFAPLGIPLSPASINQILPATSTNLFYLPYEGEISPTNTNWKONLPRDEIKLIENICWT 1210
Db 1136 IFAPLGIPLSPASINQILPATSTNLFYLPYEGEISPTNTNWKONLPRDEIKLIENICWT 1195
Qy 1211 LMDRLGYPKFMD 1222
Db 1196 LMDHLGYPKFMD 1207

RESULT 6
US-10-697-828-9
; Sequence 9, Application US/10697828
; Publication No. US20040185546A1
; GENERAL INFORMATION:
; APPLICANT: Rosen, Steven
; APPLICANT: Lee, Jin Kyu
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: Novel Glycosyl Sulfotransferases GST-4alpha, GST-4beta, & GST-6
; FILE REFERENCE: UCAL-138DIV
; CURRENT APPLICATION NUMBER: US/10/697,828
; CURRENT FILING DATE: 2003-10-29
; PRIOR APPLICATION NUMBER: 09/593,828
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: 60/144,694
; PRIOR FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 596
```

```
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-697-828-9

Query Match
Best Local Similarity 48.2%; Score 3147; DB 16; Length 596;
Matches 596; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 627 MNRNGAMVDVDAHYKMFDFHGHGSPMASIOEASQADEFKRWTFQVNVTFQMESTIT 686
DB 1 MNRNGAMVDVDAHYKMFDFHGHGSPMASIOEASQADEFKRWTFQVNVTFQMESTIT 60

QY 687 RIAYVFYGPINVSRCRFDSSNPGQLISLVNNTTHVSVIVTDYHNLKTRFNLYLFGGFG 746
DB 61 RIAYVFYGPINVSRCRFDSSNPGQLISLVNNTTHVSVIVTDYHNLKTRFNLYLFGGFG 120

QY 747 ASVADQOQITRFGLTQAIKVPVRHRIIPPFQGFKNIAVGLILCISLVILTFQWRFYLS 806
DB 121 ASVADQOQITRFGLTQAIKVPVRHRIIPPFQGFKNIAVGLILCISLVILTFQWRFYLS 180

QY 807 FRKLMRWILVLVIALWFIELLDVWSTCSQPICAKWTTRTEAGSKKSLSGEGHHMDLPDVV 866
DB 181 FRKLMRWILVLVIALWFIELLDVWSTCSQPICAKWTTRTEAGSKKSLSGEGHHMDLPDVV 240

QY 867 ITSILPGSABEILKQLFFNSDFLYIRVPTAYIDIPETELEIDSFVDACEWKVSDIRSGHF 926
DB 241 ITSILPGSABEILKQLFFNSDFLYIRVPTAYIDIPETELEIDSFVDACEWKVSDIRSGHF 300

QY 927 RLLRGMQLSVQDTKLHQLNIHLEPNRGKLAQYFAMNKDKKFKRRRESLPEQRSMKG 986
DB 301 RLLRGMQLSVQDTKLHQLNIHLEPNRGKLAQYFAMNKDKKFKRRRESLPEQRSMKG 360

QY 987 AFORDAEYIRALRHLYVYSPVLSLSSGNTLKLHFQEVLGASMRALYIVRDPRAW 1046
DB 361 AFORDAEYIRALRHLYVYSPVLSLSSGNTLKLHFQEVLGASMRALYIVRDPRAW 420

QY 1047 IYMWLNKSPSLYSLKNVPBHLAKLFKIEGKGKCNLNSGYAFEPYPLRKLKSLSKSNV 1106
DB 421 IYMWLNKSPSLYSLKNVPBHLAKLFKIEGKGKCNLNSGYAFEPYPLRKLKSLSKSNV 480

QY 1107 SLLSHLWLANTAALRINTDLPSTVOLVKFEDIVHFPQKTTIRIPAFGLIPSPASLNQ 1166
DB 481 SLLSHLWLANTAALRINTDLPSTVOLVKFEDIVHFPQKTTIRIPAFGLIPSPASLNQ 540

QY 1167 ILFATSTNLYLYEGEISPTNTNVKQNLPRDEIKLIENICWTLMDRLGYPKPM 1222
DB 541 ILFATSTNLYLYEGEISPTNTNVKQNLPRDEIKLIENICWTLMDRLGYPKPM 596

RESULT 7
US-09-984-429-71
; Sequence 71, Application US/09984429
; Publication No. US20040010132A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 53 Human Secreted Proteins
; FILE REFERENCE: P2018P2
; CURRENT APPLICATION NUMBER: US/09/984,429
; CURRENT FILING DATE: 2001-10-30
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: 60/244,591
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 09/288,143
; PRIOR FILING DATE: 1999-04-08
; PRIOR APPLICATION NUMBER: PCT/US98/21142
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 60/061,463
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 60/061,529
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 60/071,498
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 60/061,532
; NUMBER OF SEQ ID NOS: 219
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 71
; LENGTH: 126
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (126)
; OTHER INFORMATION: Xaa equals stop translation
US-10-150-111-71

Query Match
Best Local Similarity 8.7%; Score 568; DB 14; Length 126;
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 811 MRWILILVIALWFIELLDVWSTCSQPICAKWTTRTEAGSKKSLSGEGHHMDLPDVVITSL 870
DB 1 MRWILILVIALWFIELLDVWSTCSQPICAKWTTRTEAGSKKSLSGEGHHMDLPDVVITSL 60

QY 871 PGSGAEILKQLFFNSDFLYIRVPTAYIDIPETELEIDSFVDACEWK 917
DB 61 PGSGAEILKQLFFNSDFLYIRVPTAYIDIPETELEIDSFVDACEWK 107

RESULT 8
US-10-150-111-71
; Sequence 71, Application US/10150111
; Publication No. US20030078386A1
; GENERAL INFORMATION:
; APPLICANT: Rubin et al.
; TITLE OF INVENTION: Secreted Protein HPEAD48
; FILE REFERENCE: P2018PDI
; CURRENT APPLICATION NUMBER: US/10/150,111
; CURRENT FILING DATE: 2002-05-20
; PRIOR FILING DATE: 1999-04-08
; PRIOR APPLICATION NUMBER: 09/288,143
; PRIOR FILING DATE: 1999-04-08
; PRIOR APPLICATION NUMBER: PCT/US98/21142
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 60/061,463
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 60/061,529
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 60/071,498
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 60/061,527
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 60/061,536
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 60/061,532
; NUMBER OF SEQ ID NOS: 219
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 71
; LENGTH: 126
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (126)
; OTHER INFORMATION: Xaa equals stop translation
US-10-150-111-71

Query Match
Best Local Similarity 100.0%; Pred. No. 9.1e-43;
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 811 MRWILILVIALWFIELLDVWSTCSQPICAKWTTRTEAGSKKSLSGEGHHMDLPDVVITSL 870
DB 1 MRWILILVIALWFIELLDVWSTCSQPICAKWTTRTEAGSKKSLSGEGHHMDLPDVVITSL 60

QY 871 PGSGAEILKQLFFNSDFLYIRVPTAYIDIPETELEIDSFVDACEWK 917
DB 61 PGSGAEILKQLFFNSDFLYIRVPTAYIDIPETELEIDSFVDACEWK 107
```

Db 61 PGSGAEILKQLFFNSDFLYIRVPTAYIDIPETELEIDSFVDACEWK 107

RESULT 9

US-10-021-660-128
; Sequence 128, Application US/10021660
; Publication No. US20030152926A1
; GENERAL INFORMATION:
; APPLICANT: Murray, Richard
; APPLICANT: Glynn, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Waceon, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: NO. US20030152926A1el Methods of Diagnosis of Angiogenesis,
; TITLE OF INVENTION: Compositions and Methods of Screening for Angiogenesis
; TITLE OF INVENTION: Modulators
; FILE REFERENCE: 018501-000710US
; CURRENT APPLICATION NUMBER: US/10/021,660
; CURRENT FILING DATE: 2001-12-06
; PRIOR APPLICATION NUMBER: US/09/784,356
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: US 09/637,977
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 135
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 128
; LENGTH: 411
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-021-660-128

Query Match 2.3%; Score 147; DB 14; Length 411;
Best Local Similarity 18.2%; Pred. No. 0.0008;
Matches 89; Conservative 58; Mismatches 149; Indels 194; Gaps 16;

| | | |
|----|------|---|
| Qy | 813 | WILILVIALMFIEL-----LDVWSTC-----SOPICAKWTRTEAGSKSLSS 855 |
| Db | 5 | WKAVILLALASIAIQYTAIRTTAKSPHTCPGLAEGLAERLC-----EESPTAYNLRS 59 |
| Qy | 856 | EGHMDLPDVVITSLPGSGAEILKQLFFNSDFLYIRVPTAYIDIPETELEIDSFVDACE 915 |
| Db | 60 | KTH-----IILATTRSGSSFGVQLFNQHLDFVYLFEPL----- 93 |
| Qy | 916 | WKVSDIRSGHFRLLRGWLQSLVQDTKLHLQNIHLHEPNRGKLAQYFAMNKDKKFKRRE 975 |
| Db | 94 | -----YHVQNTLIPRFTQG-----KSPADRRV 115 |
| Qy | 976 | SLPEORSOMKGAFRDAEYI-----RALRRHLVYVPSARPV-----LS 1013 |
| Db | 116 | MLGASRDLLRSLYDCDLYFLENYIKPPPVNHTDRIFRGASRVLCSPVCDPPGPADLV 175 |
| Qy | 1014 | LSSGSWTLKLHFFQEVILGA-----SMRALYIVRD 1042 |
| Db | 176 | LEEGDCVRCGGLNLTVAEAACRERSHVAIKTVRVPEVNDLRALVEDPRLNLKVIQLVRD 235 |
| Qy | 1043 | PRAWIYSLMYSKPSLSYSLKNVPEHLAKLFKIEGKGKCNLNSGYAFEYEPRLKELSKS 1102 |
| Db | 236 | PRGILASRSETFRDY-----RLWRLWYGTGR-----KPNLDVTOL-TTVCEDF 279 |
| Qy | 1103 | SNVAVS--LLSHLWLANATAALRINTDLLPTSQYLVKFEDEVHFPQKTERIFAFILGIP- 1159 |
| Db | 280 | SNVSTGLMRPPLKLG-----KYMVLVRYEDLARNPMKTEEIVGVFLGIPLD 325 |
| Qy | 1160 | -----SPASLNQILPATSTNLFYLPYEGEISPTNTNWKQNLPRDEIKLIENI 1207 |
| Db | 326 | SHVARWIONNTRGDPITLGHKHYGTVRN-----SAATAEKWRFLSYDIVAFAQNA 375 |
| Qy | 1208 | CWTLMDRLGY 1217 |
| Db | 376 | COQVLAQLGY 385 |

RESULT 10

US-10-211-462-97

; Sequence 97, Application US/10211462
; Publication No. US20040033495A1

GENERAL INFORMATION:

; APPLICANT: Murray, Richard
; APPLICANT: Glynn, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Aziz, Natasha
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Angiogenesis, Compositions and
; TITLE OF INVENTION: Methods of Screening for Angiogenesis Modulators
; FILE REFERENCE: 018501-006200US
; CURRENT APPLICATION NUMBER: US/10/211,462
; CURRENT FILING DATE: 2003-02-13
; PRIOR APPLICATION NUMBER: US 09/784,356
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: US 09/791,390
; PRIOR FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: US 60/310,025
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US 60/334,244
; PRIOR FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 230
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 97
; LENGTH: 411
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-211-462-97

Query Match 2.3%; Score 147; DB 15; Length 411;

Best Local Similarity 18.2%; Pred. No. 0.0008;
Matches 89; Conservative 58; Mismatches 149; Indels 194; Gaps 16;

| | | |
|----|------|---|
| Qy | 813 | WILILVIALMFIEL-----LDVWSTC-----SOPICAKWTRTEAGSKSLSS 855 |
| Db | 5 | WKAVILLALASIAIQYTAIRTTAKSPHTCPGLAEGLAERLC-----EESPTAYNLRS 59 |
| Qy | 856 | EGHMDLPDVVITSLPGSGAEILKQLFFNSDFLYIRVPTAYIDIPETELEIDSFVDACE 915 |
| Db | 60 | KTH-----IILATTRSGSSFGVQLFNQHLDFVYLFEPL----- 93 |
| Qy | 916 | WKVSDIRSGHFRLLRGWLQSLVQDTKLHLQNIHLHEPNRGKLAQYFAMNKDKKFKRRE 975 |
| Db | 94 | -----YHVQNTLIPRFTQG-----KSPADRRV 115 |
| Qy | 976 | SLPEORSOMKGAFRDAEYI-----RALRRHLVYVPSARPV-----LS 1013 |
| Db | 116 | MLGASRDLLRSLYDCDLYFLENYIKPPPVNHTDRIFRGASRVLCSPVCDPPGPADLV 175 |
| Qy | 1014 | LSSGSWTLKLHFFQEVILGA-----SMRALYIVRD 1042 |
| Db | 176 | LEEGDCVRCGGLNLTVAEAACRERSHVAIKTVRVPEVNDLRALVEDPRLNLKVIQLVRD 235 |
| Qy | 1043 | PRAWIYSLMYSKPSLSYSLKNVPEHLAKLFKIEGKGKCNLNSGYAFEYEPRLKELSKS 1102 |
| Db | 236 | PRGILASRSETFRDY-----RLWRLWYGTGR-----KPNLDVTOL-TTVCEDF 279 |
| Qy | 1103 | SNVAVS--LLSHLWLANATAALRINTDLLPTSQYLVKFEDEVHFPQKTERIFAFILGIP- 1159 |
| Db | 280 | SNVSTGLMRPPLKLG-----KYMVLVRYEDLARNPMKTEEIVGVFLGIPLD 325 |
| Qy | 1160 | -----SPASLNQILPATSTNLFYLPYEGEISPTNTNWKQNLPRDEIKLIENI 1207 |
| Db | 326 | SHVARWIONNTRGDPITLGHKHYGTVRN-----SAATAEKWRFLSYDIVAFAQNA 375 |
| Qy | 1208 | CWTLMDRLGY 1217 |
| Db | 376 | COQVLAQLGY 385 |

RESULT 11

US-10-408-765A-395
; Sequence 395, Application US/10408765A

```

; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408.765A
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 395
; LENGTH: 411
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-395

Query Match      2.3%; Score 147; DB 16; Length 411;
Best Local Similarity 18.2%; Pred. No. 0.0008;
Matches 89; Conservative 58; Mismatches 149; Indels 194; Gaps 16;

QY      813 WILVIALWFIEL-----LDVWSTC-----SQPICKAKWTRTEAGSKSLSS 855
Db      5 WKAVLLALASIAIOYTAIRTTAKSFHTCPGLAAGLAERLC-----ESPTFAYNLSR 59
QY      856 EGHMDLPDVVITSLPGSGAEILKQLFPNSSDFLYIRVPTAYIDIPETELEIDSFVDACE 915
Db      60 KTH-----ILILATTRSGSFVQGFNQHLDFYLFEP-----93
QY      916 WKVSDIRSGHFRLLRGWLQSLVQDTKLHLQNIHLHPEPNRGKLAQYFAMNKDKRKFRR 975
Db      94 -----YHVONTLIPRFTQG-----KSPADRRV 115
QY      976 SLPQRQSMKGAFDRDAEYI-----RALRRHLVYPSARPV-----LS 1013
Db      116 MLGASRDLRLSLYDCDLYFLENYIKPPPVNHTDRIFRGASRVLCSPVCDPPGPADLV 175
QY      1014 LSSGSWTLKLHFFQEVLGA-----SMRALYIVRD 1042
Db      176 LEEGDCVRKCGLLNLTVAAEACRERSHVAIKTVRPVPEVNDLRALVEDPRLNLKVIQLVRD 235
QY      1043 PRAWIYMLYNSKPSLYSLKNVPEHLAKLFKIEGKGKCNLNSGYAFEYELPKRLSKSK 1102
Db      236 PRGILASRSETFRDTY-----RLWRLWYGTGR-----KPNYLDVTQL-TTVCEDF 279
QY      1103 SNAYS--LLSHLWLANTAALRIINTDLLPTSQYLVKFEDIVHPQKTTTERIFAFGLIPL- 1159
Db      280 SNSVSTGLMRPPLKLG-----KYMVLRYEDLARNPMKKTTEEIYGFGLIPLD 325
QY      1160 -----SPASLNQILFATSTNLFYLPYEGEISPTNTVWKNQNLPRDEIKLIENI 1207
Db      326 SHVARWQNTNRGDPFLGKHKGTYVRN-----SAATAEKWRFRLSYDIVAFAQNA 375
QY      1208 CWTLMDRIGY 1217
Db      376 CQVLAQLGY 385

RESULT 12
US-10-723-860-1544
; Sequence 1544, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnick, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators

```

```

; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723.860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1544
; LENGTH: 411
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-723-860-1544

Query Match      2.3%; Score 147; DB 16; Length 411;
Best Local Similarity 18.2%; Pred. No. 0.0008;
Matches 89; Conservative 58; Mismatches 149; Indels 194; Gaps 16;

QY      813 WILVIALWFIEL-----LDVWSTC-----SQPICKAKWTRTEAGSKSLSS 855
Db      5 WKAVLLALASIAIOYTAIRTTAKSFHTCPGLAAGLAERLC-----ESPTFAYNLSR 59
QY      856 EGHMDLPDVVITSLPGSGAEILKQLFPNSSDFLYIRVPTAYIDIPETELEIDSFVDACE 915
Db      60 KTH-----ILILATTRSGSFVQGFNQHLDFYLFEP-----93
QY      916 WKVSDIRSGHFRLLRGWLQSLVQDTKLHLQNIHLHPEPNRGKLAQYFAMNKDKRKFRR 975
Db      94 -----YHVONTLIPRFTQG-----KSPADRRV 115
QY      976 SLPQRQSMKGAFDRDAEYI-----RALRRHLVYPSARPV-----LS 1013
Db      116 MLGASRDLRLSLYDCDLYFLENYIKPPPVNHTDRIFRGASRVLCSPVCDPPGPADLV 175
QY      1014 LSSGSWTLKLHFFQEVLGA-----SMRALYIVRD 1042
Db      176 LEEGDCVRKCGLLNLTVAAEACRERSHVAIKTVRPVPEVNDLRALVEDPRLNLKVIQLVRD 235
QY      1043 PRAWIYMLYNSKPSLYSLKNVPEHLAKLFKIEGKGKCNLNSGYAFEYELPKRLSKSK 1102
Db      236 PRGILASRSETFRDTY-----RLWRLWYGTGR-----KPNYLDVTQL-TTVCEDF 279
QY      1103 SNAYS--LLSHLWLANTAALRIINTDLLPTSQYLVKFEDIVHPQKTTTERIFAFGLIPL- 1159
Db      280 SNSVSTGLMRPPLKLG-----KYMVLRYEDLARNPMKKTTEEIYGFGLIPLD 325
QY      1160 -----SPASLNQILFATSTNLFYLPYEGEISPTNTVWKNQNLPRDEIKLIENI 1207
Db      326 SHVARWQNTNRGDPFLGKHKGTYVRN-----SAATAEKWRFRLSYDIVAFAQNA 375
QY      1208 CWTLMDRIGY 1217
Db      376 CQVLAQLGY 385

RESULT 13
US-09-817-514A-8
; Sequence 8, Application US/09817514A
; Patent No. US20020078478A1
; GENERAL INFORMATION:
; APPLICANT: French-Constant, Richard
; APPLICANT: Bowen, David
; APPLICANT: Rocheleau, Thomas
; APPLICANT: Waterfield, Nicholas
; TITLE OF INVENTION: DNA SEQUENCES FROM PHOTORHABDUS LUMINESCENS
; FILE REFERENCE: 61645
; CURRENT APPLICATION NUMBER: US/09/817,514A
; CURRENT FILING DATE: 2000-03-26
; PRIOR APPLICATION NUMBER: US 60/191806
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 2504

```


Prior Application Data:
; APPLICATION NUMBER: US 60/007,255
; FILING DATE: 06-NOV-1995
Prior Application Data:
; APPLICATION NUMBER: US 08/608,423
; FILING DATE: 28-FEB-1996
Prior Application Data:
; APPLICATION NUMBER: US 08/705,484
; FILING DATE: 28-AUG-1996
ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.93804
TELECOMMUNICATION INFORMATION:
; TELEPHONE: 608-251-5000
; TELEFAX: 608-251-9166
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2504 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-10-262-794A-12

Query Match
Best Local Similarity 18.1%; Pred. No. 0.024;
Matches 231; Conservative 180; Mismatches 443; Indels 422; Gaps 60;

QY 29 FSTFEESVSNSEWA-----VFTDDIDQFKTKQVDFRPNQKLKSMHLPSLYFD 78
DB 1003 FTDWERYNKRYSTWAGSELVYYPENYVPTQRIQGTQKMD-ALLQSIQSQLNADTVED 1061
QY 79 AGEIQAMQKRSRASHLHLFAIRSAVTV-----MLSNPTYLPPPKA-----D 122
DB 1062 AFKTY-LTSFEQVANLKVISAHONVNVDQGLTYFIGIDQAAPGYIYWRVSDHSCENGK 1120
QY 123 FAA-----KWNIEYGNLPLALYCLLCPEDKVAPEFVLEYMDRMVGYKDWLNAAPGDEV 178
DB 1121 FAANAWGEMKI-----TCVNPWKNIIRPVV--YMSRL--YLLWLEQSKSD- 1165
QY 179 PIGSLTGFATAPDFLNLNDRHQKYLEKIW-----VITEMEYYSKVRSGWKQLLH 232
DB 1166 -----DGKTTIYO--YNLKLHTR-----YDGSWNTPTFDVTEKVKNYTSSTDAESEL-- 1212
QY 233 NHQATNMIALTGALVGTGVDGSKANIQKAVVDVMEKTMFLNHIYDGLSDEGVAAGSY 292
DB 1213 -----GLYCTGYQG-----EDTLLVMFYSMQSS-----YSSY 1239
QY 293 T---AKSVTQVFLAQRHFNINLND--NNMLKMHFWFYATLPGFQRTVGIADSNYNWF 347
DB 1240 TDNNAPVTGLYIFADMSDDNWTNAQATYNNNSY-----PQFTVMDPDS----- 1286
QY 348 YGPESQLVFLDKFLKNGAGNLAAQQRKHPRKDGMPVPSTAQKRWSTLHTEYIYDPQLT 407
DB 1287 -----NKKVITRRVNRVYAEVDE-----IPSSV-----TSNSYNSWGDHSLT 1323
QY 408 POPPADYGTAKIHFTFPWGVVYTYGAGLPN-----TQNTFVS-FKSGKLGRAV 455
DB 1324 -----MLYGSVENITFESAEDLRLSTNMALSIITHNGYAGTRRI 1363
QY 456 YDIVHFQPSYWDGRGPNPHEPDQNSFTFAPNGOVFVSEALYGPKLHLNVLNVPAP 515
DB 1364 QCNLMKQYASLGDFIYDSSFD--DANRNLVP--LF-----KFGDENSDDSIICYNE 1414
QY 516 SPSSQCNKPWBGQLGECAQMLKWTGE-----EVGDAAGEIITASQHGEMVFSGEAVSAYS 571
DB 1415 NPSSE-DKKWYFSSKDDNKTDYNGGTCIDAGTSNKDFYVNLQIEIVISVTGGYSSYK 1473
QY 572 SAMRLKSVYRALLLNLSQTLVVDHIERQEDSPINSVSAFHNLDIDFKYIPKFMNRYN 631
DB 1474 IS-----NTGIDSAKV--KVTYKAG 1496
QY 632 GAMMDVMDAHYKMFDFDHGNSPWASIQEABQAAEFKKRWTFQVNVTFQWEST--ITRIA 699

DB 1497 G-----DDQIFTADNSTVYPQ-----QAPSPFEEMIYQFNNTLIDCKNLNFDNOA 1542
QY 690 YVYFGPYINVSSCRIDSSNPGLOISLVNNTNTEHVVSIVTDYHNL-----KTRFNVL 741
DB 1543 HIEIDFTATAQDGRFLGAEFTFIIPVTYKVLGTENYALYSENNGVOYMOIGAYRRLNTL 1602
QY 742 GFEGFASVADQG-----QITRFGLGTQAIKVPVRHRIIFPFPGFKNIAVGLI 789
DB 1603 FAQQLVSRANRIGIDAVLSMETQNIQEPQLGAGTYVQLVLDKYDESHGNTKNSFALEYVDI 1662
QY 790 LCISLVILTFQWRFYLSFKLMRWILILVIALWFIELL-----DVMSTCSQPICAKW---- 841
DB 1663 FKENDSFVIYQGLSETSTQVVK-----VFLSYFIEATGNKNHLW-----VRKYOKET 1711
QY 842 -----TRTEAE-----GSKSLSSSEGHMDLPDVTITSLPGSGA---- 875
DB 1712 TDKILFDRTDEKDPHGWFSLDDHKFTFGLSSAQAALKNDSEPMDF-----SGANAL 1761
QY 876 -----EILKOLFNFSSD--FLYIRVPTAVIDIPETETELEIDSVPDACEWKV 918
DB 1762 YFWELFYPTPMMAHRLLEQNFDAANHFWRVWSPSGYI-----VDGKIAYHNVV 1813
QY 919 SDIRSGHFLRLRGWLQSLVQDTKLJHLQNTIHLHPN-----RGKLAQYFA-----MN 964
DB 1814 -----RPLEEDTSMNAQQLDSTDPDAVQDDPMHYKVATFMATLDLLMAR 1858
QY 965 KKKKPKFKRESLPQRSONMKGAFDRDAEYIARL-----RHLVYYPSPARPVLSLSSGS 1018
DB 1859 GDAAYRQLERDLAEAKMW-----YTQALNLLGDPEQVMLSTTWTANPTLGNAAASK 1908
QY 1019 WTLKLHFFOEVLGASMRALYIIVRDPRAWIYSMLYNSKPSLSYSLKNVPEHLAKLFKIEGK 1078
DB 1909 TTQQVR--QQVL--TQRLNSRVKTP-----LLGTANSLTAL-FLPQENSKL-----K 1951
QY 1079 G-----KCNLNSGYAFEPYELRKLKELSKSNVAVLSLHMLWLANTAALRINTDLLPT 1130
DB 1952 GYWRTLAQRMFNLRNLSIDGQPLSLPYAKADPKALLSAAVSASQGA----- 2001
QY 1131 SYQLVKFEDIVH-FPQ 1145
DB 2002 --DLPKAPLTIHRFPQ 2015

RESULT 15
US-10-754-115-59
; Sequence 59, Application US/10754115
; Publication No. US20040208907A1
; GENERAL INFORMATION:
; APPLICANT: Hey, Timothy
; APPLICANT: Schieper, Amanda
; APPLICANT: Bevan, Scott
; APPLICANT: Bintrim, Scott
; APPLICANT: Mitchell, Jon
; APPLICANT: Li, Ze Sheng
; APPLICANT: Ni, Weiting
; APPLICANT: Zhu, Baolong
; APPLICANT: Merlo, Don
; APPLICANT: Apel-Birkhold, Patricia
; APPLICANT: Meade, Thomas
; TITLE OF INVENTION: Mixing and Matching TC Proteins for Pest Control
; FILE REFERENCE: DAS-104XCI
; CURRENT APPLICATION NUMBER: US/10754,115
; CURRENT FILING DATE: 2004-01-07
; PRIOR APPLICATION NUMBER: US 60/441,723
; PRIOR FILING DATE: 2003-01-21
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 59
; LENGTH: 2504
; TYPE: PRT
; ORGANISM: Photorhabdus luminescens
US-10-754-115-59

| Query Match | 2.2% | Score 145; | DB 16; | Length 2504; |
|-----------------------|-------------------|---|-------------|--------------|
| Best Local Similarity | 18.1%; | Pred. No. 0.024; | | |
| Matches 231; | Conservative 180; | Mismatches 443; | Indels 422; | Gaps 60; |
| QY | 29 | FSTFEESVSNYSWA-----VFTDIDQFKTKQVDFRPNOKLKKSLMLHPSLYPD | 78 | |
| DB | 1003 | FTDWERNKRYSTWAGVSELVYPENYVDPTRQIGQTKMD-ALLQSINQSLNADTVED | 1061 | |
| QY | 79 | AGEIQAMQRKSRASHLHFRAIRSAVTV-----MLSNPTYIYPPPKHA-----D | 122 | |
| DB | 1062 | APKTY-LTSEQVANLKVISAYHDNVNDQGLYFFIGIDQAAEGTYWRSVDHSCKENG | 1120 | |
| QY | 123 | FAA-----KMWIEYGNLPPALALYCLCPDKVAFEFVLEYMDRMVGYKMWLENAPGDEV | 178 | |
| DB | 1121 | FAANAWGEMNKI-----TCAVNPWKNIIRPVV--YMSL--YLLMLEQQSKSD-1165 | | |
| QY | 179 | PIGHSLTGPFATPDFLYNLLDNHRKYLEKIW-----VITEEMEYSKVRSMGKQLLH | 232 | |
| DB | 1166 | -----DGKTTIYO--YNLKLAIHR--YDGSWNTPFTFVTEKVKNYTSSSTDAEASL-- | 1212 | |
| QY | 233 | NHQATNMIALLTGALVTGVDKGSKANIWKQAVVDVMEKTMFLNLHIVDGSLDEGVAYGSY | 292 | |
| DB | 1213 | -----GLYCTGYOG-----EDTLILVMFYMSQSS-----YSSY | 1239 | |
| QY | 293 | T--AKSVTQYVFLAQRFHFINLLD--NNWLKMHFMFYATATLLPGFQRTVGIADSNYNWF | 347 | |
| DB | 1240 | TDNNAPTVGLYIFADMSDDMTNAQATNYNNSY-----PQPDVMAADPDS----- | 1286 | |
| QY | 348 | YGPESQLVLDKILKMGAGNLAAQIRKURPKDGPWSPSTAORWSTLHTEYIWDPOLT | 407 | |
| DB | 1287 | -----NKKVITRRVNNRYAEDYE-----IPSSV-----TSNSNYSMGDHSLT | 1323 | |
| QY | 408 | POPPADVGTAKIHTFPNWGVVTVYGALPN-----TOTNTFVS-FKSGKLGRGV | 455 | |
| DB | 1324 | -----MLYGGSVNTPFSSAEDBLSTNMAUSIHHGVAGTRRI | 1363 | |
| QY | 456 | YDIVHFQPYGWIWDGWSFNPGEHPDQNSFTFAPNGQVVFSEALYGPKLHLNNLVFAP | 515 | |
| DB | 1364 | QCMLMKQYASLGDKFIYYDSSF--DANRNLVP--LF-----KFGKDENSDDSIYNE | 1414 | |
| QY | 516 | SPSSQCNKPWEGQGECAQMLKWTE-----EVGDAGEIITASQHGEMVTFVSEANVAYS | 571 | |
| DB | 1415 | NPSSE-DKKWYFSSKDDNKTDADYNGGTQCIDAGTSNKFDFYNNLQIEIIVSVTGGYMSYK | 1473 | |
| QY | 572 | SAMRLKSVYRALLLNSQTLVVDHIERQDSESPINSVSAFPHLLDIDFKYIPKFMNRYN | 631 | |
| DB | 1474 | IS-----NPLNI-----NTGDISAKV--KVTVKAG | 1496 | |
| QY | 632 | GAMMDVMDAHYKNMFDHNGSNPMASIOEAQAAEFKKRWTFQVWVTFQMEST--ITRIA | 689 | |
| DB | 1497 | G-----DDQIFADNSYVPO-----QPAPSFEMIYQFNLLIDCKNLNFDNQ | 1542 | |
| QY | 690 | YVFYGPYINSSCRFIDSNSPGLQISLVNNTNTEHVSVITDYHNL-----KTRFNYL | 741 | |
| DB | 1543 | HIEIDFTAQDGRFLGAETFIIPVTKVLGTENVIALYSENNGVQYMQICGAVRTRLNTL | 1602 | |
| QY | 742 | GFGGFASVADQ-----QITRFGLGTQAIKVPVRHRIIPFPGFKFNIAVGLI | 789 | |
| DB | 1603 | FAQQLVSRANRGIDAVLSMETQNTQEPQLGAGTYVQLVDKIDSHSHTGNTKNSFAIEYVDI | 1662 | |
| QY | 790 | LCISLVILTTFQWRPYLGFRLMRWILILVIALWPIELL-----DWMSTCSQPICAKW--- | 841 | |
| DB | 1663 | FKENDSFVIYQGEISETSQTVK-----VFLSYFIEATGNKNHL-----VRAKYOKET | 1711 | |
| QY | 842 | -----TRTEAB-----GSKKSLSEGGHMDLPDVVITSLPGSGA--- | 875 | |
| DB | 1712 | TDKILFORTDEKDPHGWFSLDDHKTFSGLSSAAQAKNDSEBMDF-----SGANAL | 1761 | |
| QY | 876 | -----EILKQLFNSSD--FLYIRVPTAYTIDIPETELEIDSFVDACAEKW | 918 | |
| DB | 1762 | YFWELFYTTMMWHLRLQEQNFDAANWFRYVMSPGYI-----VDGKTAIYHMNV | 1813 | |

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2005, 08:33:38 ; Search time 27.7171 Seconds.
(without alignments)
4242.027 Million cell updates/sec

Title: US-10-697-828-15
Perfect score: 6532
Sequence: 1 MPKGGAPPWIMALMFTGHL.....LIENICWTLMRLGYPKPMF 1222
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Maximum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_79:.*
1: PIR1.*
2: PIR2.*
3: PIR3.*
4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----------|---------------------|
| 1 | 251.5 | 3.9 | 672 | 2 T36083 | hypothetical prote |
| 2 | 192 | 2.9 | 776 | 2 D38354 | oligo alginate lya |
| 3 | 192 | 2.9 | 776 | 2 AC2928 | oligo alginate lya |
| 4 | 133 | 2.0 | 484 | 2 J50261 | N-acetylglucosamin |
| 5 | 127 | 1.9 | 1456 | 1 A36563 | mannose receptor p |
| 6 | 125 | 1.9 | 4344 | 1 A53489 | dynein heavy chain |
| 7 | 123 | 1.9 | 1044 | 2 B72338 | galactosyltransfer |
| 8 | 122 | 1.9 | 1747 | 2 T43162 | vitellogenin - gyp |
| 9 | 120.5 | 1.8 | 944 | 2 D82926 | hypothetical prote |
| 10 | 113.5 | 1.7 | 484 | 2 JC7350 | N-acetylglucosamin |
| 11 | 113.5 | 1.7 | 3744 | 2 S46715 | hypothetical prote |
| 12 | 113 | 1.7 | 2216 | 2 S78398 | hypothetical prote |
| 13 | 112.5 | 1.7 | 4199 | 2 S76412 | hypothetical prote |
| 14 | 111 | 1.7 | 921 | 2 A51698 | isoleucyl-tRNA syn |
| 15 | 110.5 | 1.7 | 1248 | 2 C89874 | autolysin [importe |
| 16 | 110.5 | 1.7 | 1455 | 1 A48925 | mannose receptor p |
| 17 | 110 | 1.7 | 542 | 2 S54260 | bromodomain protei |
| 18 | 110 | 1.7 | 2095 | 2 S29529 | genome polyprotein |
| 19 | 110 | 1.7 | 3655 | 2 T38084 | TRAP-like protein |
| 20 | 109.5 | 1.7 | 1441 | 2 T00335 | hypothetical prote |
| 21 | 109 | 1.7 | 2401 | 2 T28676 | riophtry protein - |
| 22 | 108.5 | 1.7 | 866 | 2 T06454 | probable lipoxigen |
| 23 | 108.5 | 1.7 | 936 | 2 H71862 | probable cytochrom |
| 24 | 108.5 | 1.7 | 1215 | 2 T43916 | chitinase A [limpor |
| 25 | 108 | 1.7 | 1102 | 2 S44772 | C29E4.4 protein - |
| 26 | 108 | 1.7 | 1589 | 2 T22668 | hypothetical prote |
| 27 | 108 | 1.7 | 2076 | 2 S15999 | fatty-acyl-CoA syn |
| 28 | 107 | 1.6 | 771 | 2 H84845 | hypothetical prote |
| 29 | 107 | 1.6 | 1131 | 2 F82875 | hypothetical prote |

| | | | | | |
|----|-------|-----|------|----------|---------------------|
| 30 | 107 | 1.6 | 1816 | 2 A84845 | probable ABC trans |
| 31 | 107 | 1.6 | 1969 | 2 T38495 | hypothetical prote |
| 32 | 106.5 | 1.6 | 532 | 2 H96840 | hypothetical prote |
| 33 | 106.5 | 1.6 | 2619 | 2 T24588 | hypothetical prote |
| 34 | 106 | 1.6 | 486 | 2 JC7351 | N-acetylglucosamin |
| 35 | 106 | 1.6 | 921 | 2 AC1327 | isoleucyl-tRNA syn |
| 36 | 105.5 | 1.6 | 518 | 2 B83472 | hypothetical prote |
| 37 | 105.5 | 1.6 | 906 | 1 IJXLC2 | N-cadherin 2 precu |
| 38 | 105 | 1.6 | 386 | 2 S58755 | ribosomal protein |
| 39 | 105 | 1.6 | 877 | 1 IJBOCN | N-cadherin precurs |
| 40 | 105 | 1.6 | 910 | 2 T22050 | hypothetical prote |
| 41 | 105 | 1.6 | 924 | 2 T22044 | hypothetical prote |
| 42 | 104.5 | 1.6 | 586 | 2 D90250 | glycogen synthase |
| 43 | 104.5 | 1.6 | 664 | 2 D82056 | ABC transporter, A |
| 44 | 104.5 | 1.6 | 686 | 1 ALDYAT | amylase A (EC 3.2. |
| 45 | 104.5 | 1.6 | 929 | 2 T17392 | .vrlL protein - Dic |

ALIGNMENTS

RESULT 1

T36083

hypothetical protein SCE134.01c - Streptomyces coelicolor (fragment)

C:Species: Streptomyces coelicolor

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 12-Jul-2004

C:Accession: T36083

R:Saunders, D.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, April 1999

A:Reference number: Z21596

A:Accession: T36083

A:Status: preliminary; translated from GE/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-672 <SAU>

A:Cross-references: UNIPROT:Q8CJW4; EMBL:AL049661; PIDN:CAB41199.1; GSPDB:GN00070; SCOEDB:AL049661

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCOEDB:SCE134.01c

C:Superfamily: oligosaccharide lyase

Query Match 3.9%; Score 251.5; DB 2; Length 672;
Best Local Similarity 22.6%; Pred. No. 9.5e-10;
Matches 125; Conservative 86; Mismatches 249; Indels 93; Gaps 27;

| | | | |
|----|-----|--|-----|
| Qy | 80 | GEIQMRQKSRAGHLHLFRAIRSAVTVMLSNPTVYL---PPPKHADFAAKWNEYGNLNP | 136 |
| Db | 6 | GRDLSLRGDLNGTHAAQWRRLHEQC-----DWYRQNPTEHPEASITY---FGPAAA | 55 |
| Qy | 137 | PLALYCLLCPEDKVAFEVLEVMYDMRVGYKDWLVENAPGVDEVPIGHSLTGPATPFDLYN | 196 |
| Db | 56 | NLALAYRLTQQ-RGYLBEAARWISTCVAYPHWGRAHMPDHDLDAGWLLHGLSLAYSWLGE | 114 |
| Qy | 197 | LLNHRROKYLEKIWIWTEMEYSKV---RSWCKQLLHHQATNMIALLTGALVTGVDK | 253 |
| Db | 115 | DLEPERREILRAKLEQGERLHSAFETTGWMSSAYWNH---NWIC-WTGATAGYAL | 170 |
| Qy | 254 | GSRANIWKQAVDVMKMTMLLNHIVDGLDEGVAGSYTAKSVTVQVFLAQRH----- | 307 |
| Db | 171 | GRSE--WTKAARANLETULTMLPE--DGSSEGVVYWRVGVFWLAIHTDLVQREQADLM | 226 |
| Qy | 308 | ----FNINLNDNLWLMKMFYFATLPGFQRTVGIADSNYNWYFSPESQLVFLDKFLK | 363 |
| Db | 227 | STGGFLRNT--TRW-RLH-----QSAPGFEENIDHGDC-HDRSGHVALYRLASAYQ | 276 |
| Qy | 364 | NGAGNWLAAQOI-RKH-----RPKDGPMVSTAQRMWSTLHTEIWDVDPQLTP-OPP | 411 |
| Db | 277 | DGTAQMLGNLVAERHFWREAYESGVRPGVMP-----EAFLELLWYDPRVTPAAPD | 326 |
| Qy | 412 | ADVCTAKIHTFPNMGVVVYAGLPNTQTNFVSPKSGKLGRAVYDVIVHFQFYSWIDGWR | 471 |
| Db | 327 | REPOTA---YFPDUGQITARTGMDSAA--TCVSPKAAFGGHRADWDGH--RLKAAGWD | 379 |
| Qy | 472 | SFNPGEHBPQNSFTFAPNGQVFVSEALYG--PKLSHLNNVLVFPAPSPSQCNKFWEQQL | 529 |

Db 380 AMSAGHHHPDAGAFVLHSHGAFLAVDSGYSNHKEAAHNLVLV-----DGEWGA 428
Qy 530 GECAQWLKWTG--BEVGDAGEIITASCHGEMVFSGEAVSAYSSAMRLKSVYRALLLN 587
Db 429 DE-GRYHVYEGIPERRARVDVL--AQDG-FAHATAESAAMFSERLGVQRVDRTLVWTP 484
Qy 588 SQTLLVVDHIERQ 600
Db 485 LGRVVILDELEAE 497

RESULT 2
D98354
oligo alginate lyase (AB011415) [imported] - Agrobacterium tumefaciens (strain C58, Cere
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 12-Jul-2004
C:Accession: D98354
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-776 <KUR>
A:Cross-references: UNIPROT:Q8UBJ1; GB:AE007870; PIDN:AAK90358.1; PID:g15160399; GSPDB:G
C:Genetics:
A:Gene: AGR_L3558
A:Map position: linear chromosome
C:Superfamily: oligosaccharide lyase

Query Match 2.9%; Score 192; DB 2; Length 776;
Best Local Similarity 19.1%; Pred. No. 2.2e-05;
Matches 137; Conservative 94; Mismatches 279; Indels 208; Gaps 31;
Qy 33 EESVSNYSWAVFTDDIDQFKTKQVDFRPNQKLLKSM-----HPSLYFDA 79
Db 99 QKSATAHNSWT-----VRSFEISALPKTLPGRSARHAAQAOTSHPLWLNS 146
Qy 80 GEIQAMRQ--KSRASHLHLFRAIRSATVMSLNPYYLPP--PKHADFAAKNNEIYGNL 135
Db 147 EQLSAFADAVAKOPNHCWGAEFEYKSEVPWLERPVMPPEQYPNNTVATLWQMY---- 202
Qy 136 PPLALYCLLCPEDKVAEFVLEYMDRMVGYKD-----ENAPGDE 177
Db 203 -----IDCQEVYAIRH-LAIAGRVLGRDLDLDSARKWLLAAVAWDTKGATSRAYNDE 254
Qy 178 VPIGHSLTGATAFDFLYNLL--DNHR--RQVLEKIWVITEEMYEYSKVRSGKQLLHN 233
Db 255 AGF-RVVVALAWGYDMLYDHLSEDERTRVSVLLERTREVDHVIHARIHVFP--YDS 310
Qy 234 HQATNMIALLTGALVTGVDKSGKANIWKQAVVDVMEKTMFLNHHIVDGLSDE-----G 286
Db 311 HAVRSLSAVLTPACIALQGESDEAGELDYTFEFL-ATLSPWAGTDGGWAEGPHYWMTG 369
Qy 287 VAYGSYTAQSVQYV--FLAQHFNNINLNNLWQHFWFYATLPGFORTVGIADSNY 344
Db 370 MAYLIEANLIRSITYGYDLYQRP-----FQNGRFLYTKAPGTRRAN- 413
Qy 345 NWFYGPESQVLFDKF-----ILKNGAGNWLAAQOIRKHPKDGPMVPSTAQRWS 393
Db 414 ---FGDDSTLGDLPGLKGLYVNRQFAGVTGNHGYQWYFDHIK-----ADATGTE 459
Qy 394 TLHTEYIWD-----POLTPQPPADYGTAKIHTFPNNGVVVYAGLPTQTNT 441
Db 460 MAFYNGWDLNFDLVRHDYDQVEAVSPADLPALAVFDDIGWATIQKMDPEDRHLQ- 518
Qy 442 FVSPKSKLGRVAVDVIHFQPSWIDGWSRPNPGHEHPQNSFTFAPNGQVFSBALY 501
Db 519 FV-FKS-----SPYGS-----SHSHGDQNAFVLYAHGEDLAIQSGY- 554

Qy 502 PKLSHLNNLVFAPSPSQCNKPWEGQ-----LGCEAOMLKWTEGVEGDAAGEIITA 553
Db 555 -----YVAF-----NSQHLNWRQTRSKNAVILGGKGQYAEKDKALARAAGRIVSV 602
Qy 554 SQHGMVFSGEAVSAYSSAMRL-KSVYRALLLLNSQTLVVDHIERQ----- 601
Db 603 BEQPGHVRIVGDATAAYQVANPLVQKVLRETHFVDSYFVIVDEVECSPEQLWLCHTL 662
Qy 602 DSPINSVSAPFFNLIDIDFKYIYPKPNRYNGAMMDVDAHYKMFDFDHGNSPMASIQ 659
Db 663 GAPQTRSSSF-----RYNGRKA-----GFYQGFVYSGGTPOISAVE 699

RESULT 3
AC2928
oligo alginate lyase [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 12-Jul-2004
C:Accession: AC2928
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.;
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClellan,
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, E
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AC2928
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-776 <KUR>
A:Cross-references: UNIPROT:Q8UBJ1; GB:AE008689; PIDN:AAAL43841.1; PID:g17741383; GSPDB:G
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu3025
A:Map position: linear chromosome
C:Superfamily: oligosaccharide lyase

Query Match 2.9%; Score 192; DB 2; Length 776;
Best Local Similarity 19.1%; Pred. No. 2.2e-05;
Matches 137; Conservative 94; Mismatches 279; Indels 208; Gaps 31;
Qy 33 EESVSNYSWAVFTDDIDQFKTKQVDFRPNQKLLKSM-----HPSLYFDA 79
Db 99 QKSATAHNSWT-----VRSFEISALPKTLPGRSARHAAQAOTSHPLWLNS 146
Qy 80 GEIQAMRQ--KSRASHLHLFRAIRSATVMSLNPYYLPP--PKHADFAAKNNEIYGNL 135
Db 147 EQLSAFADAVAKOPNHCWGAEFEYKSEVPWLERPVMPPEQYPNNTVATLWQMY---- 202
Qy 136 PPLALYCLLCPEDKVAEFVLEYMDRMVGYKD-----ENAPGDE 177
Db 203 -----IDCQEVYAIRH-LAIAGRVLGRDLDLDSARKWLLAAVAWDTKGATSRAYNDE 254
Qy 178 VPIGHSLTGATAFDFLYNLL--DNHR--RQVLEKIWVITEEMYEYSKVRSGKQLLHN 233
Db 255 AGF-RVVVALAWGYDMLYDHLSEDERTRVSVLLERTREVDHVIHARIHVFP--YDS 310
Qy 234 HQATNMIALLTGALVTGVDKSGKANIWKQAVVDVMEKTMFLNHHIVDGLSDE-----G 286
Db 311 HAVRSLSAVLTPACIALQGESDEAGELDYTFEFL-ATLSPWAGTDGGWAEGPHYWMTG 369
Qy 287 VAYGSYTAQSVQYV--FLAQHFNNINLNNLWQHFWFYATLPGFORTVGIADSNY 344
Db 370 MAYLIEANLIRSITYGYDLYQRP-----FQNGRFLYTKAPGTRRAN- 413
Qy 345 NWFYGPESQVLFDKF-----ILKNGAGNWLAAQOIRKHPKDGPMVPSTAQRWS 393
Db 414 ---FGDDSTLGDLPGLKGLYVNRQFAGVTGNHGYQWYFDHIK-----ADATGTE 459
Qy 394 TLHTEYIWD-----POLTPQPPADYGTAKIHTFPNNGVVVYAGLPTQTNT 441
Db 460 MAFYNGWDLNFDLVRHDYDQVEAVSPADLPALAVFDDIGWATIQKMDPEDRHLQ- 518

Qy 442 FVGFKSKGLGRAVYDIYVHFQPSYWDGWSRNFSGHEHPDQNSFTFAPNGQVVFSEALY 501
 Db 519 FV-FKS-----SPYGL-----SHSGDQNAFVLAHGEDLAIQSGY- 554
 Qy 502 PKLSHLNVLVFPAPSPQCNKPWEGQ-----LGECAQWLKWTGEEVGDAGEIITA 553
 Db 555 -----YVAF-----NSQHLNWRQTRSKNAVLIGKGQYAEKQKALARRAAGRIVS 602
 Qy 554 SQHCEMVFVSGEAVSSAMRL- KSVYRALLLNSQTLVVDHIERQE----- 601
 Db 603 EQPGHVRIVGDATAAIVQANPLVQKVLRETHFVNDYFVIVDEVECSBQELQWLCHTL 662
 Qy 602 DSPINSVAFPHNLIDIFKIPYKFNRYNGAMMDVDAHYKMFWDHGHGNSPMASIQ 659
 Db 663 GAPQTGRSSF-----RYNGRKA-----GFYGVQPVSSGGTTPQISAVE 699

RESULT 4
 JE0261
 N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) - human
 C:Species: Homo sapiens (man)
 C:Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 01-Mar-2002
 C:Accession: JE0261
 R:Uchimura, K.; Muramatsu, H.; Kaname, T.; Ogawa, H.; Yamakawa, T.; Fan, Q.W.; Mitsuoka, T.
 J. Biochem. 124, 670-678, 1998
 A:Title: Human N-Acetylglucosamine-6-O-sulfotransferase involved in the biosynthesis of
 A:Reference number: JE0261; MUID:98391845; PMID:9722682
 A:Accession: JE0261
 A:Molecule type: mRNA
 A:Residues: 1-484 <UCH>
 A:Cross-references: DDBJ:AB014679
 C:Comment: This protein catalyzes the transfer of sulfur of sulfate from 3'-phosphoadenosine
 C:Superfamily: chondroitin 6-sulfotransferase
 C:Keywords: sulfotransferase

Query Match 2.0%; Score 133; DB 2; Length 484;
 Best Local Similarity 26.2%; Pred. No. 0.18;
 Matches 53; Conservative 36; Mismatches 81; Indels 32; Gaps 9;

Qy 1034 MRALYIVRDPRAWYSLMYNSKPSLY--SLKNVPE-----HLAKLFKIEGK-GKCNLS 1085
 Db 279 LKVIHLVDRPA-VASSRIRSRHGLRESLQVVRSDPRAHMPPFLAAGHLGAKKEGV 337
 Qy 1086 GYAFYEPPLRKLKSKNSVALLSHLWLANAALINTDLPSTVOLVKFEDIVHFPQ 1145
 Db 338 GGPADYHAL-----GAMEVICNSMAKTLQALQ-PPDWLQGHVLYVRYEDLVGEPV 387
 Qy 1146 KTTERTIAFLGIPLSA-----SLNQILPATSNLPYLYPEGEISPTN-----TNVWKQNL 1197
 Db 388 KTLRRVYDFVGLLVSPMEQFALNWTSGSGSSKPFV-----VSARNATQANAWRTALT 442

Qy 1198 RDEIKLIENICWTLMRLGYPK 1219
 Db 443 FQKIQVEEFYCPMAVLGYER 464

RESULT 5
 A36563
 mannose receptor precursor - human
 C:Species: Homo sapiens (man)
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
 C:Accession: A36563; A60926; A44255; B44255; C44255; D44255; F44255; G44255; H44255
 R:Taylor, M.E.; Conary, J.T.; Lennartz, M.R.; Stahl, P.D.; Drickamer, K.
 J. Biol. Chem. 265, 12156-12162, 1990
 A:Title: Primary structure of the mannose receptor contains multiple motifs resembling
 A:Reference number: A36563; MUID:90324192; PMID:2373685
 A:Accession: A36563
 A:Molecule type: mRNA
 A:Residues: 1-1456 <TAY>
 A:Cross-references: UNIPROT:P22897; GB:J05550; NID:g188675; PIDN:AAA59868.1; PID:g188676

A:Note: parts of this sequence, including the amino end of the mature protein, were confi
 R;Ezekowitz, R.A.B.; Sastry, K.; Bailey, P.; Warner, A.
 J. Exp. Med. 172, 1785-1794, 1990
 A:Title: Molecular characterization of the human macrophage mannose receptor: demonstrati
 A:Reference number: A60926; MUID:91079783; PMID:2258707
 A:Accession: A60926
 A>Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-1333,'T',1335-1456 <EZE>
 A:Cross-references: GB:X55635
 A:Note: translation of the nucleotide sequence is incomplete
 A:Note: in the authors' translation additional residues Pro-Glu-Ile are shown after 497-1
 R;Kim, S.J.; Ruiz, N.; Bezouska, K.; Drickamer, K.
 Genomics 14, 721-727, 1992
 A:Title: Organization of the gene encoding the human macrophage mannose receptor (MRC1).
 A:Reference number: A44255; MUID:93052405; PMID:1294118
 A:Accession: A44255
 A>Status: preliminary; nucleic acid sequence not shown; not compared with conceptual trar
 A:Molecule type: DNA
 A:Residues: 155-233,'KSA',238-283;346-428;492-569;631-714;716-719;783-820,'N',822-865;'
 A:Note: sequence extracted from NCBI backbone (NCBIP:118415, NCBIP:118421, NCBIP:118428,
 C:Genetics:
 A:Gene: GDB:MRC1
 A:Cross-references: GDB:133759; OMIM:153618
 A:Map position: 10p13-10p13
 C:Superfamily: phospholipase A2 receptor; C-type lectin homology; fibronectin type II re
 C:Keywords: duplication; lectin; tandem repeat; transmembrane protein
 F:1-18/Domain: signal sequence #status predicted <SIG>
 F:168-209/Domain: fibronectin type II repeat homology <2FI>
 F:223-340/Domain: C-type lectin homology <LCH1>
 F:362-486/Domain: C-type lectin homology <LCH2>
 F:945-1079/Domain: C-type lectin homology <LCH3>

Query Match 1.9%; Score 127; DB 1; Length 1456;
 Best Local Similarity 18.7%; Pred. No. 2.7; Mismatches 284; Indels 218; Gaps 36;
 Matches 138; Conservative 97;

Qy 180 IGHSLTGFATAPDFLYN-----LLDNHRRQKYLEKIWVITEEMYEY---SKVRSWG--- 227
 Db 518 IGHSLTGFATAPDFLYN-----LLDNHRRQKYLEKIWVITEEMYEY---SKVRSWG--- 227
 Qy 228 -----KQLLNHHQNTMIALLTG--ALVTGVDKSGKANIWQAVVDVMEKWTFLNHHVVDG 281
 Db 578 WTIEEVRFTHMNDMPGRKPGCVAMRTIAGG---LWD--VLKCDKAKFVCKHWAEG 631
 Qy 282 -----SLDEGVAYGYSYAKSVTQYVFLAQRHF-----NIN 311
 Db 632 VTHPPKPTTTPPKCPEDWGASRTSLCFKLYAKGHEKKTWTFESRDFCRALGGLASIN 691
 Qy 312 NLNN-----WL-----KWH--FW--FYATLLPGFQRTVGIADSNYNYWFGPESQLVLD 358
 Db 692 NKEEQTIWRLITASGSYHKLFWLGLTYGSPSEGFTWSDGSPVSYENWAYGEPNNYQNV 751
 Qy 359 -----KFLKNGAGNWLAAQIRKUR-----PKDGPWPSTAOARWSTLH 396
 Db 752 YCGELKGDPTMSWINDINCEHLNNWIC-QIQGQTPKPEPTPAQDNP--PVTEGW-VIY 807
 Qy 397 TEYIWDPLTPQPPADYGTAKIHTFENMGVVTYAGLPLNTQTNTVFSKSGKLGRAVY 456
 Db 808 KDQYVY-----FSKEKETMDNARAFCKRNFGDL---VSIQSESEKFLWKYVNRNDAGAY 860
 Qy 457 DIVHF-----QPSYWDG-----WRSFNPCHHPDQNSFTFAPNG-----QVF 494
 Db 861 FIGLLISLKKFAWMDGSKVDYVSWATGEPNPNANEDENCVTWYNSGFWNDINGYNPAP 920
 Qy 495 VSEALYGPKLSHLN--NVLVFAVSPSQCNKPWEGQGECAQWLKWTGEE----- 542
 Db 921 ICQR-----HNSINATTVMPTMPSVPSGCKEGWNFYKNCFKIFGFMEERKNWQEAR 976
 Qy 543 -VGDAAGEITASQHGEMVFSVGEAVSAYSSAMPLKSVYRALLLNLSQTLVVDHIERQE 601
 Db 977 CIGFG-----GNLVSIQNEKEQAFUTYHMKDSTFSAWTGLND----- 1013

Db 295 HI-----NHKKDPAWMLQIIGMLRKIDERYTLHAGDFQ-----BIRYANYPKH----- 338
Qy 701 SCRFIDSSNPGLOISLNVN--TEHVSIVTDYHNLKTRFNVLGFG-GFASVADQGQITRF 758
Db 339 ---FI--KDAGLEKXVLYGWEDVNAFLDKVYLLSTSHESGYNIAEMAKG----- 388
Qy 759 GLGTAIVKVRHD-----RIIFPGFKFNIAVGLICISLVILTFQWRFY----- 804
Db 389 -----IKPIHNYAGAKTQWPDLLVFN-----IDEVIRVTSRDYNSEKYSFVE 434
Qy 805 --LSFRKLMWILLVIALFIELLDVWSTCSQPICAKWTREABEGSKSLSSGHHMDL 862
Db 435 KNCLEKQITSIIV-----QLQDNRTKNSIHKST--TDTENSFAKIKEYSKID- 486
Qy 863 PDVVTSLPSGG-----AEILKOLF-FNSDFLVRPTAY-----IDIPE 902
Db 487 SFTIMDLPGKLSRSEFVSLLERFIIINKAKILEVGTGAFSIELALREADVTGIDIDP 546
Qy 903 TELEI-----DSFVDACEWKVSDIRSGHFRLLR-----GWLQSLVQDTKLH 943
Db 547 TSIELAIRISKDYNVENVEPKVD-----GPKLTSEPKQEPDIAFNMGVVEHFKDDDIK 602
Qy 944 LQNTLHLPNRKLAQVAFAMNKKRKFRESLPEORSQMGK-----FORDABYIRALR 999
Db 603 MLK-----QMGVAKFVVGVPYSGSFVYKTA--KETAQKLGAWEGYGFERDFTLEPLI 654
Qy 1000 RHLVYVPSARPVLSSSGSWTLKLFHQEVLGASMRALYIVRDPRAVIYMLNYSKPSLY 1059
Db 655 RRAGLIP-----LH-----EEVIGVLAEPFYLRRINPEW----- 684
Qy 1060 SLKXVPEHLAKLFKIE--GGKGNKLNLS--GYAEYEPPL--RKELSKSKNAVALLS 1110
Db 685 PLK--IAENLQKYFOGEKVGWNLICFAIKWPCYADEFLKLDHKKI--KFESQISLIT 739

RESULT 8
T43162
vitellogenin - gypsy moth
C:Species: Lymantria dispar (gypsy moth)
C>Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jul-2004
C:Accession: T43162; T43268
R:Hiremath, S.T.; Lehtoma, K.
A:Title: Structure of the gypsy moth vitellogenin gene.
A:Reference number: 222378; MUID:97468497; PMID:9327584
A:Description: Gypsy moth vitellogenin mRNA.
A:Reference number: 222321
A:Accession: T43162
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1747 <HIR>
A:Cross-references: UNIPROT:Q25269; EMBL:U60186; NID:g1399940; PID:g1399941; PIDN:AAB033
R:Hiremath, S.; Lehtoma, K.
Arch. Insect Biochem. Physiol. 36, 151-164, 1997
A:Title: Structure of the gypsy moth vitellogenin gene.
A:Reference number: 222378; MUID:97468497; PMID:9327584
A:Accession: T43268
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-819, 'O', 821-1747 <H12>
A:Cross-references: EMBL:U90756; NID:G2894826; PID:g2894827; PIDN:AAC02818.1
C:Genetics:
A:Introns: 11/1; 725/2; 796/3; 1120/1; 1410/1; 1477/2
C:Superfamily: boll weevil vitellogenin

Query Match 1.9%; Score 122; DB 2; Length 1747;
Best Local Similarity 20.4%; Pred. No. 8.2;
Matches 268; Conservative 174; Mismatches 536; Indels 336; Gaps 70;

Qy 36 VSNYSEWAVFTDDIDQ-----FKTKVQDFRPNQ---KUKSKMLH-----PS 74
Db 459 IAKVDMWFIFRDVIAQTQAPQFOLIEWTKTIQGFLLAEVSSIAKTIRHLTKOTMK 518
Qy 75 LYFDAGEI-QAMRQKSRASHLHLFRATRSVTVMLSNPTY-----YLPKPPKHADFAAKWN 128

Db 519 QFFELALCPVLEBQ-----YLNSSALIAAABFIRLGSIDYEHSPYPPYVHGFRPPLKND 574
Qy 129 E-YIGNNLPLLA-LYCLLCPEDKVAPEFVLEVMYDMVMGYKMDLVNAPGDEVPIGHSU-- 184
Db 575 RFIIENILPRLSQLLKQAVEQEDYGKTLVIVKAIGTLGHREILKVFTPYLGEGQIKVSTYL 634
Qy 185 -TGFAFAFDLYNLLDNHRRQKYLEKIWIITEWYKSVRSWKQLLH-NHQATNMIAL 242
Db 635 RVQMILALKPLAEQKORYVRSALFSILMNTABPYE---VRVAAAMILFLNEPTTDLRV 690
Qy 243 LTGALVTGVDKSGKANIKOAVDVMEKTMFLN-----HIVDG-----SLDEGVA 288
Db 691 M--AQLTND--PSIHRAVLKSSITETAATLEGPKYWHLAKAAQSVKELVTSDEFQYA 744
Qy 289 YG-----SYTAKSVTOYVFLAQRHENINNDNMLKMHF---W---FYVATLLPGPQR 335
Db 745 YSAWPAKPFDEEVTRY-----RLANYIGSDGLIPNLSLTWNNRPFYGR-----GAEN 795
Qy 336 TVGIADSNYMWYGPESQLVFLDKFILKNGAGNWLAAQIRK-HRPKQGPMPV-STAORWS 393
Db 796 TVGFTSNVQDIFDYIKQLILQPR--KSTTEHQFSADKIAKLNIKQDPLQPIESSIYK 853
Qy 394 TLHTEYIWDYDPLTPQPPADYGTAKIHTFPNAGV--TYGAGLPNTQNTTFVSFKSKL-- 450
Db 854 IPDKPEFW-----PLDQPNIKQLVNDVFGVKELENGVQSHYTKVFYSKQSVIFP 904
Qy 451 ---GGRAYD---IVHFOYSWIDGWRSFNPGHEHPDQ-----SFTFAPNQVQVSE 497
Db 905 IATGEPFVYKPPVHVHLO--TKLSKINYPPTSKDISSNIQTEIOFTFARN-----ME 957
Qy 498 ALXGPKLSHLNVLVFAFSPSCQNKPEW-----GQLGECAQWLKWTGEEVGDAAGEITA 553
Db 958 GSVGFDITNTQLANIGVVKYQINLPVKINVIGOSGKFVKL-----EPLHT 1005
Qy 554 SQHGEVMVFSGEAVSAYSSAMRLKSVYRALLNLSQTLVVDHIEROEDSPINSVAFPH 613
Db 1006 DQDLILLHSHWSPYSAYQQTNTTTPSHDLTKLIKELKVAENKELGQOQFNLL----- 1060
Qy 614 NLDDIDFKYIPYKPMNRVNGA-----MMVDMDAHVKMFDFDHGNSPMASI 658
Db 1061 -----FKYHGFSSNTYRNAGDFVQDADLDFSDPMSLNDLAETHYE---FTYSGK----- 1106
Qy 659 QEAQEAAPFKRWI-----QFVNVY-----FQMESITIRIAYVFGYINV 699
Db 1107 QSKTKAVTTAAVDVSHNLSTGKFGAHRFSDVTPNSAPRREIVKRS-----SGIKT 1161
Qy 700 SSCRFID--SSNPGLOISLNVNTEHVSVITVDYHNLKTRFN-YLFGGPGFASVADQGIT 756
Db 1162 ATARVIDFSASFEGLO-KLEYAVATAAVAGSMVD---LKTQPAVFMG-----SOSDNGOI- 1211
Qy 757 RFLGTOAIKVPVRHDIRIIFPGFKFNIAVGLICISLVILTFQWRFVLSFR---KLNR 812
Db 1212 -----NAVFKLOKQPMADLDFHKLNSAVKYLEAD---VTYGENSININFGHTERSQE 1262
Qy 813 WILITVIALWFIELLDVWSTCSQPICAKWTREABEGSKSLSS-----EGHMDLPDVI 867
Db 1263 YAEOLKNSLWA-----NQCAQ-----ENAAQNKFGQCHNLVIESHAPDRFKASI 1307
Qy 868 T--SLPGSGABILKQLFFN--SSDFLYIRVPTAVIDIPETELED-SFVD-----ACBWK 917
Db 1308 TYKHIPAAHTALLDSYIQLGWSKGFY--NPSKRLPVQGLEANASVVDQTANVAVTWM 1365
Qy 918 VSDIRSGH-----RLRGWLQSLVQDTKHLON---IHLHPNRGKLAQYFAMNKDK 967
Db 1366 NGQVRFNLPNTVITPALTAYQDYGIEDSWTHFPANSYSYHQYEP-----FCTVDTGT 1418
Qy 968 KRKFRR-----SLPE-----QRSQMK-----GAFDRDAEYIR 996
Db 1419 VKTFSDYDYNVTLPEIWTVLHQAQTNWELVLVLAKEPNEAKKAEIKGKGLDYISHKT 1478
Qy 997 ALRHL--VYFSPARPVLSSSGSWTLKLFHQEVLGASMRALYIVRDPRAVIYMLNYSK 1055

Db 1479 ATGKHLEVNIP-----YSAANNKANVKVETNAQIADVADGLTTY--WDDVAETPLQVSNH 1531
Qy 1056 PSLSYLNKVPHEHLAKLPIKGG-----GKCNLNSGVAFYEPYELRKLSKSN 1104
Db 1532 PDRVLLHLSDGLHLPLDGRGIFTTSQYRNITRGICGQNSG-----DPL--DDYKTPILG 1584
Qy 1105 AVSLLSHLWNTAAALRINTDLPTSQVLVKFEDI-----VHPPOKTERIPAF 1154
Db 1585 IVDHSQHFGAFT-----LDLEKNSIQQWKKAQETAYQPKLTHVTILRF 1631

RESULT 9
D82926
hypothetical protein UUI166 [imported] - Ureaplasma urealyticum
C:Species: Ureaplasma urealyticum
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: D82926
R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.
submitted to GenBank, February 2000
A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a min
A:Reference number: A82870
A:Accession: D82926
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-944 <GLA>
A:Cross-references: GB:AE002116; GB:AF222894; NID:g6899118; PIDN:AAF30573.1; GSPDB:GN001
A:Experimental source: serovar 3; biovar 1
C:Genetics:
A:Gene: UUI166
A:Genetic code: SGC3

Query Match 1.8%; Score 120.5; DB 2; Length 944;
Best Local Similarity 17.5%; Pred. No. 4;
Matches 212; Conservative 148; Mismatches 388; Indels 461; Gaps 58;

Qy 44 VFTDDIDQFKTKVQD-----FRPNQKLKSLMLPSLYFDAGEIQAMRKSRASHLHLFR 98
Db 9 IYLDLEGYSQKNIEUSIDSLIEPNQK-----PIHLCPFR 43
Qy 99 AIRSAVTMLNSNTYLYLPPPKHADPAKNEIYGNLPLALYCLLCPBDBK---VAPEFV 155
Db 44 ILKGFKITNTGQSLLVNDK--DFSTVMN-IKASNEPIIAYW-----DKSFTYNFKGS 94
Qy 156 LEYMDRMVGVKDWLVENAPGDEVPIGHSLTGFATADFLYNLLDNHRRQKYLEKIWIITE 215
Db 95 LNVISK-----DYARIDPSNTLKNF-----KPYYPQAWV--- 125
Qy 216 EMVEYSKVRSGKQLLHNHQAQTNMIALLTGALVTGVDKSGKANIWKQAVVDVMEKTMFL 275
Db 126 --QNYQANNVSTIVNDKEHTWIKTITVTLIVKYPVVEDEIYQNE----- 171
Qy 276 NHAVDGLS-DEGVAYGSYTAQSVQYVF-LAQRHFNINLNDNNLWKHFWFYATLLPGF 333
Db 172 ---QGNLVSHKAKFKQWTTMDLKEYLFIIPDSHAQANQ-----WLMVNEVIPFY 217
Qy 334 QRTVGTADSNYNWFGPESQVLVDFDKFLKNGAGNWLAAQOIRKRPKDGPMVPSTAQHRS 393
Db 218 STT-----SKRYE-----GAQITITNVNDLTTSSGNDILATIQCPQGDGYL-----WP 261
Qy 394 TLAHEYIWDYDQ-----LTPQPPA-----DYGTAKIHTFPNMGVVTYGA 432
Db 262 VIQNGKLTDERKEKPLGVKSIQLLTNGPYALNSMLITGTNNYNDTRLSANKELLYKA 321
Qy 433 GLP-----NTQNTNFTVSKGKLGGRAYVDIVHFQPYSWIDGWSRPNPGEHPDQNS 484
Db 322 KIPVASKFFNVSAPTSNSVYLYNAFPEIEKIYK-----DQYEKQKD 362
Qy 485 FTEAPNGQVFSVSEALYCPKLSH-----LNNVLVAPSPSSQCNKPWE----- 526
Db 363 FTFDHNAITKINGGLL---LNHKWDASNNALSY--BDNRYPFKLWDEGPKIHASFNFKL 417
Qy 527 -----GOLGBCAOWLKWLTGTEV--GDAAGBIITASQHGEMVFSV--GEAVS----- 568

Db 418 CGTGAQKGGSLTENAEQKKPLADAVVTQGVAGEVVST-----VPESIGRNVSQGFNLNN 471
Qy 569 -----AYSSAMRLKSVYRALLLNSQTL-----LVVDHLIERQEDSPINSVSAP 611
Db 472 LGLWSFNMGFAKTKAEQIKBATK--IGLNNITMPGPTNLHTFLNH-----NAP 517
Qy 612 PHNLIDIDPKYIPYKPMN--RYN-GAMMDVWDADHYKMFWDHGHGNSPMASIOEAEQAAE-F 667
Db 518 MYDL--IKYLPYSYKETIROQLGELID-----NTPLSLGKVRGIVDIL 559
Qy 668 KKRQTQFVNVTFQMBESTITRIAYFYGPYINVSRCRFDISSNPNLQIQLISLVNNTHTVWSI 727
Db 560 YPGWTKVEDATGYNMSNI-----WLNTRGFENFI---APCLPLELEVEN-----KL 601
Qy 728 VTDYHNL-----KTRPNVLYGFG-----FASVADQ-----GOIT- 756
Db 602 KADLNLNIDIFTNKPACT-FSPILGIPAIACCVHFDITDRFAKFEVDDELKVFDTLIGQTTY 660
Qy 757 RFGLGTQAIKVPVRDRII-----FPFGFK--FNIAVG-----LILCISL- 794
Db 661 RDTYIRKDKKPLDYWDSFCEALPPNTPLSYEPYAVINIASGKREDYNRVKYPVIDITDIK 720
Qy 795 -----VILTFQWRFYLSFRKMRWILILVIALWFIELLDVMTSCQ-----PICAKWT 842
Db 721 GLGKCDIKLTAPSGFYSPDARKD-----EFKNVYEAIFESNGKYTDDMSQWS 768
Qy 843 RTEAGSKSLSSSEGHMDLPDVVITSLPGSGAEILKQLFNSSDFLYIRVPTAVIDIPE 902
Db 769 NYINFSTLDEINSVGTLSYPK-----PPVETDFLKNTKNSDDGVNASLQT----- 815
Qy 903 TELEIDSFVDACEMKVSDIRSGHFRLLRGWLQSLVQDTKLHLQNLHLHHPNKGKLAQVPA 962
Db 816 -----INETLYKVD-----NSLLSARQYKRPFRFYFEPDSSKLANLEFG 854
Qy 963 MNKDKRKFKRRESLPEQRQSMKGAFDRDAEYIRALRRHLVYYPSPARVPLSSSGSWTLK 1022
Db 855 Y-----GKGCN 1082
Qy 1077 G---GKGCN 1082
Db 918 NWIQKGTWN 926

RESULT 10
JC7350
N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) [validated] - mouse
C:Species: Mus musculus (house mouse)
C:Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 09-Jul-2004
C:Accession: JC7350
R:Uchimura, K.; Fasaki, F.; Kadamatsu, K.; Matsuoka, T.; Yamakawa, N.;
Biochem. Biophys. Res. Commun. 274, 291-296, 2000
A:Title: Diversity of N-acetylglucosamine-6-O-sulfotransferases: Molecular cloning of a
A:Reference number: JC7350; MUID:20374462; PMID:10913333
A:Accession: JC7350
A:Molecule type: mRNA
A:Residues: 1-484 <UCH>
C:Cross-references: UNIPROT:O99NB0; UNIPROT:O9EP78; DDBJ:AB040710
C:Comment: This enzyme, having a type II transmembrane topology and N-linked glycosylatic
sulfate linkage formation. This enzyme is involved in synthesis of L-selectin ligand in
C:Keywords: glycolysis; sulfate transport; sulfotransferase; transmembrane protein

Query Match 1.7%; Score 113.5; DB 2; Length 484;
Best Local Similarity 18.8%; Pred. No. 4.5;
Matches 81; Conservative 72; Mismatches 171; Indels 107; Gaps 18;

Qy 848 GSKSLSEGHM--DLPDVVITSLPGSGAEILKQLFNSSDFLYIRVPTAVIDIPEL 905
Db 83 GSPGNLSAAGEAVTQEKQHIYVHATWRTGSSGLFELFNQHPDVFYLYBPMWHL----- 135

Qy 906 EIDSFVDACBWKVSDIRSGHFRLLRGWLQSLVOD---TKLHLQNIHHEPNRGKLAQYFA 962
Db 136 -----WQA--LYPGDAESLQAGALRDLWLSLFRCDFSVLRLYAQPGDGERAPDSA 183
Qy 963 MNKDKKPKRR-----ESLPEQSQMKGAEDRDAE-----YIRALRRHLVY 1005
Db 184 -NUTAMLFWRWNKVCSPPLCPAARAKRVDGLVEDKACESTCPPVSLRALEAECKRY 242
Qy 1006 PSA--RPVLSLSSGWSLTKLHFFOEVL-----GASMRALYIVRDPRAIYSLMYLSNKP 1056
Db 243 PVVVIKDVRLDLG-----VLVPLLRDPGLNKKVQLPRDPA-VHNSRLKSRQ 290
Qy 1057 SL-----YSLKNVPBHLAKLFKEGKCKNINSYAFEBYELRKLKSKSNVALL 1109
Db 291 GLLRESTQVLRQRGDHFRVLLAHG-----VDARPGQARALPSPAPRDFLTSALEVI 346
Qy 1110 SHLWLANTAALRINTLLPT-----SYQLVKFEDIVHFPQKTERIFAFGLIPLS 1160
Db 347 CEAWL-----RDLFTRGAPAWLRRYRLRYEDLVQQAQLRLLRPSGL-RT 395
Qy 1161 PASLNQILFATSTNLFYLPVEGEISPTN-----TNWKNQLPRDEIKLIENICWTL 1211
Db 396 LAALDAFAPNTRGSAY-----GADRPFLSARDAREAVHVRRLSQEVRQVETACAPA 451
Qy 1212 MDRLGYPKFM 1222
Db 452 MRLLAYPRSGD 462

RESULT 11
S46715
hypothetical protein YHR099w - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein H9332.1
C;Species: Saccharomyces cerevisiae
C;Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 09-Jul-2004
C;Accession: S46715
R;Vaadin, M.
submitted to the EMBL Data Library, May 1994
A;Description: The sequence of S. cerevisiae cosmid 9332.
A;Reference number: S46715
A;Molecule type: DNA
A;Residues: 1-3744 <VAU>
A;Cross-references: UNIPROT:P38811; EMBL:U00060; NID:9487928; PIDN:AAB68923.1; PID:94879
C;Genetics:
A;Gene: SGD:TRAI
A;Cross-references: SGD:S0001141; MIPS:YHR099w
A;Map position: 8R

Query Match 1.7%; Score 113.5; DB 2; Length 3744;
Best Local Similarity 18.6%; Pred. No. 1.1e+02;
Matches 256; Conservative 196; Mismatches 482; Indels 439; Gaps 66;

Qy 26 MFAPSTFEE-----SVSNSEWAVFTDDIDQPKTKQVDFRP-----NOKLKKSMHLPSL 75
Db 218 MFSFKILSECPITWTVLYSSV-----KQLTSTSLPEFTPLIMNLLNIQIKQQO----- 265
Qy 76 YFDAGEIQMRQKSRASHLHLFRAIRSAVTWLSNPTY---YLPPEKHADE----- 123
Db 266 --EARE---QAESRGEH-----FTSISTEINRPAYCDPILAQIKATSFAYVFIRGY 313
Qy 124 AAKNNEYGNLPLALCYLL-CPEDKVAPEFVLEYMDRWV---GYK-----DWLVEN 172
Db 314 APEFLQDVNVPDIIIRLLQDCSELSARKELHATHRLSTNYKKLFLPKLDYLF- 372
Qy 173 APGDEVPIGHSLT-----GFATAPDFLYLLNDNHRQKYLEKIWIITEEMEYS--- 221
Db 373 ---ERILIGNGFTWHETLRPLAYSTVADFHNIRSELQSLSEIKTIKIYGYLLDESIAL 429
Qy 222 KVRSGKQLLHNHOATNMIALLTGALVTGVDDXGKANIWQAVVDVMEKTM---FLNHI 278
Db 430 TVQIMSAKLLN-----LVERILKLGKNGPOEAPRAKLLMIIDSYNRPKTLNRQ 481

Qy 279 VDGSLEGVAYGYSYAKSVTQYVFLAQRHFNINNNLWLNKMHFWFYATLLPGFQRTVG 338
Db 482 YDTIMK---YIGRYETHKKEAEKJK-----NSIQDN----- 510
Qy 339 IADSNYWFVGPRESQVFLDKFKILKNGAGNLAQOIRKHRPKDQPMVPSTAQRWSTLHTE 398
Db 511 -----DKESE--EFMRK-VLEFSDDDHLMPQPKKEDINDSP----- 543
Qy 399 YIWYDQLTPQPPADYGTAKIHTFPNMGVVYTGAGLPNTQTN-----TFVSFKS 447
Db 544 ----DVEMTESDKVKVNDVEMFPIKNYAPILL---LP-TPTNDPIKDAFYLYRLTMSFLK 595
Qy 448 GKUGRAVDIVIHFPQYSWIDGWRSPNPGHEHDQNSFTFAPNGQVPSVSAALYCPKLSHL 507
Db 596 ----TIHDLKVFNP-----PPNEYTVANPKLWASVSRVFSYE----- 629
Qy 508 NNVLVAFAPSPSSQCNKPEWQGLCECAQWLKWTGEEVGDAAAGEIITASQHEGMVPSGEAV 567
Db 630 -EVIVP-----KDLFHECIIGLKFFKHNEKLSPE--TTKKHFD-ISMPSLPV 673
Qy 568 SAYSSAMRLKSVYRALLLNLSQTLVVDHIERQ-----EDSPINSVSFAFFHNLDI- 617
Db 674 SATKDARELMD-YLAFMFMQMDNATFNEIIEQELPFVYVERMLEDSGLLHVAQSFLTSEIT 732
Qy 618 --DFKIPIPKMRYNGAMMDVMDAHYKMFWDHGHGNSPMASIOEAQAAEFKRWTFQV 675
Db 733 SPNFAGILLRFL---KGKLDLG----- 752
Qy 676 NYTFQMESTITR---TAYVFYGPYINVSCEFIDSSNPGLOISLVNNTTEHVHVSIVTDYH 732
Db 753 NVDFTNSVNLIRLFLKLSFMSVNLFPNINEVVLPHLNDLILNSUKYSTTAEPLVY--FY 810
Qy 733 NLKTRFNLYGFGGF---ASVADQGOITRFLGLGTQAIKVPVRHDRIIFPFGKFNIAVGL 788
Db 811 LIITLFRSIOGGFENLYRSIKELIQLVQLSQNLQMLTARLPHEREY-----V 859
Qy 789 ILCISLVILTFQWRFYLSFRKLMRWLILVIAL-WFIEL-----LDVWSTCSQPICAKWTR 843
Db 860 ELGITVPVRLSVLAPYLPF-LMK---PLVFALQOYVDLVSQGLRTLELCIDNLNTASYFD 914
Qy 844 TEAGSKSLSSGHHMDLPDVVITSLPGSGAELTKOLFNFSSDFLYRVTAYIDPIE- 902
Db 915 PIIEPVIDDVSKALFNLLQPPFNHAIHNVVRILGKLGGRNQFL--KPPT---DLTEK 969
Qy 903 TELEIDSFVDACBWKVSDIRSGHFRLLRGWLQSLVQDTKLHLQNIHHE----- 951
Db 970 TELDIDAID---FKINGMPEDVPLSVTPGQISALNLOQYSKSDIHYRKSAYKYLTCVLL 1026
Qy 952 -----PNRGKLAQYFAMNKDK-----KXKFKRRRESL-----P 978
Db 1027 LMTKSSAEFTPTNVTETLLKTAVNSIKLERIGIEKNFDLEPTVNRKDSYQENLFLRLLESV 1086
Qy 979 EQSQMKGAEDRDAEYIRALRRHLVYPSARPVLSL--SGSWTLKLHPFQEVILGASM-- 1034
Db 1087 FYATSIKELKDDAMDLLNLLDHFCLLQVNTTLNKNYNGTFFNIDKPNFMIDSSLL 1146
Qy 1035 -----RALYI--VRDPRAIYSLMYLSNKPSTLY-----SLKNVPEHLAKLF----- 1072
Db 1147 DAIPLFALSYIYIPEVREVGVLAYKRIYEKSLIYGEELALSHSFIPE-LAKQFIHLCYDET 1205
Qy 1073 --KIEG-----KGCNLSNGYAFEBYELRKLKSKSNVALLSH 1111
Db 1206 YNKRGGVGLIKVILDNVKS SVFLKKYQYNLANGLLF-----VLKDTQSEAPS----- 1254
Qy 1112 LWLANTAALRINTDLLPTSYQLVKFEDIVHFPQKTERIFAFGLIPLSPASLNQILPAT 1171
Db 1255 ---AITDSAEKLLIDLISITFADVKEED-----LG-----NKLVENT 1288
Qy 1172 STNLFYLPYEGESPTNTNVMKONLPRDEIKLIENICWTLMDRLGYD--KFMD 1222
Db 1289 LTDIVC-----EUSNANPKV--RNACOKSLHTISNL-----TGIFIVKLM 1327

| Query Match | 1.7k; | Score 113; | DB 2; | Length 2216; |
|-----------------------|-------------------|--|-------------|--------------|
| Best Local Similarity | 17.66; | Pred. No. 52; | | |
| Matches 234; | Conservative 187; | Mismatches 423; | Indels 488; | Gaps 66; |
| QY | 42 | WAVFTDDIDQKTKQKQVDFPNOKLKKMLHPS-LYFDAGEIQAMRQKRSASHLH--LF- | 97 | |
| DB | 408 | YSVISNEDSEVHTLINO--REIKPLKSIIPDPFSFLOTEATEIESDQLQKRSGVSSILFT | 465 | |
| QY | 98 | -----RAIRSAVTVMLSNPTYYLPPPKHADFAAKWNEIY-GNNLPLAL--YCLLCP | 146 | |
| DB | 466 | EHEKQMINHMLPEBIEEFIGNPTRLV---HSPLSDRWSELHSGNSPTERSTRODKLLKK | 523 | |
| QY | 147 | EDKVAF-----EFV-----LEYMDRMVGK-----DMLVENAPGDEVPIGHSL | 184 | |
| DB | 522 | QODLSVFSRSENKELVNLIKITYLKNVTISHPISDPGCGDGLVKDEP--DMDSSNKI | 579 | |
| QY | 185 | TGF-ATAFDLYNLDDNHR-----QKLEKIWV---ITEEMEYKSVRS | 225 | |
| DB | 580 | SVFNKNTIYILFHLPHDNRVGYTLNLHDDFEERFQEKADLFILSITEPDLVYHKGS | 639 | |
| QY | 226 | WGKQLLNHQATNMIALTLGALVTGVDKGSKANIMKQAVVDVMEKT-----MFL | 275 | |
| DB | 640 | FS---IYMDQKQKVVFPASNIMEAVNQSR---FIRNMIKIQSYTYGIRNVLHRPF | 692 | |
| QY | 276 | NHIVDGSLDGVA-----GSYTAQSVTYVFLAQRHF-NINLNDNLKQHFVYYA | 327 | |
| DB | 693 | NR-----SDHNLEYEIKRQDKDTLHRTIIRYMINOHLNFKKSONKWFNPILF--- | 743 | |
| QY | 328 | TLLPGFQRTGVIADSNYNWFGPESQLVFLDKILKNGAGNMLAQIRKHRPKDGMVPS | 387 | |
| DB | 744 | -----FSRTSRSVRN-----PDAY-----RYKRSNGSNFLE----- | 771 | |
| QY | 388 | TAQRWSTLHTEYIWDYDQLTPQPPADYGTAKIHTFENWGVVTYGAGLPNTQNTFVSPKS | 447 | |
| DB | 772 | -----HLEH-----FVSEQK | 781 | |
| QY | 448 | GKLGRVYDIVHQPSYWDGHSFNPGEHPDQNSFTFAPNGQVFVSEAL-----YG | 501 | |
| DB | 782 | SHFKVIQVFDLIRNQYS-ID-WSAFITDKDLSKPLRF--FLSKLLFLLNSLPPFCVSVFG | 838 | |
| QY | 502 | PKLSHLNNLVF-APSPSSONKPFEGQCGEQAQNLKWTGEEVGDGAAGEIITASOHGMV | 560 | |

A;Cross-references: UNIPROT:P74440; EMBL:D90915; GB:AB001339; NID:g1653604; PIDN:BAAL1854
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C;Superfamily: Synechocystis hypothetical protein slr0408

Query Match: 1.7%; Score 112.5; DB 2; Length 4199;
Best Local Similarity 20.9%; Pred. No. 1.5e+02;
Matches 94; Conservative 42; Mismatches 131; Indels 183; Gaps 25;
224 RSMGKQLLH-NHQATNMIALLTGALVTGVDGSKANI-----WKOAV 264
362 QSMGKLPALNELAVNNTSGNPQIAGL--CGKGGEIYNGSTWVNGPYQGGWRSAL 3119
265 VDVMEKTMFLNLHIVDGLDE--GVAYGSYT-----AKSVTQYVFLAQRH 307
3120 -----TQMAVQMGEDGSPSIUVGLADGAVIYNTQSGWRTINNFGKSVTQLSVQVQEA 3173
308 PNIN--NLNN-----WLKMH--FWFYATLLPGFQRT-----VGIADS 342
3174 SNPNIVVGLDNSEVQYQSGNGVMTQPHDDGMVYPVQQL-AVQWTSNDAQPLVVVGLGDD 3232
343 NVN---WFY---GPESQLVFLDKFKLNGAGNLAQIQRKRPKDGMPVSTAKRSTL 395
3233 NGNGSVVYQSGEGGQWTFSG--LPSGAA-----IAQMAVQWNF- 3272
396 HTEYIVDPQLTPPPADYGTAKIHTFENWGVVTYGAGLPNTQTNTFVSKSGKLGRV 455
3273 -----SSPNPNVNDLKI-----VV-----GQADSTVSYNG----- 3301
456 YDIHFQPSYIDWNR-----SNPGHEHPDQNSFT--FAPNQGVFVSEALYQPKLSHLNV 510
3302 -----NGWTATPAINSSLIQIPTLNAITVQWSANGQPQITVGLGDPEDYD--NGQ 3347
511 LVFAPSPS-----SOCKNPEGLGECAG-----WLKWT 539
3348 LWYLPNPSQSQWELQSGVNYASPIQIDSSWTESLVPSQTDNLVSYVFFGSDFNQDTV 3407
540 GE-----EVGDAAGEIITASQHGEMVFSVSG 564
3408 GTIGDDVMVGSATGESFLAGGQDQILTKG 3437

RESULT 14
AE1698
isoleucyl-tRNA synthetase [imported] - Listeria innocua (strain Clip11262)
C;Species: Listeria innocua
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C;Accession: AE1698
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecher, R.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Krett, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Maok, C.; Schluter, T.; Smoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AE1698
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-921 <GLA>
A;Cross-references: UNIPROT:Q92926; GB:AL592022; PIDN:CAC97357.1; PID:g16414641; GSPDB:Q
A;Experimental source: strain Clip11262
C;Genetics:
A;Gene: ileS
C;Superfamily: isoleucine-tRNA ligase

Query Match
Best Local Similarity 1.7%; Score 111; DB 2; Length 921;
Matches 174; Conservative 112; Mismatches 306; Indels 334; Gaps 46;
117 PPKHADFAAKNEIYGNLPLALYCLICPEKVAFFVLEWDMVGYKQWLVENAP-- 174
21 PNKEPEWQAKW-----BEEKLYEKIQE---KNAGRKAYILHDGPPY 58

175 -GDEVPIGH-----SLTGFATAPDFLNLNDNHR--OKYLEKIWIITEEM 217
59 ANGELMHGHALNKTIKDIIIVRYKSMAGFSPPY---VPGWDTGLPIETAIKAGKVKRKEM 115
218 -----YEYSKVRSMGKQ-----LLHNHOATNMIALLTGALVTG 250
116 SIAEFRKLCABAYAMTQVQDRTGKRLGINDGNDHENPITLLPEYEAQIKVFGEMAKKGY 175
251 VDKGSKANIKW-----QAVVDVMEKT---MFLNLHIVDGS--LDEGVAYGSYTAKS 297
176 IYKGGKPVYSPSESALAEIEYQDTSASIFVAEKVTDGKGLDDEGTINVIWT---T 232
298 TQYVFLAQRHFNIN-NLDNNLWLMHFWFYA--TLLPGFORTVGIADSNVNFYGPESQL 354
233 TPMTIPANMIGITVNPDLDDYVVISAGEKYVVAEALLPSLREKLG----- 276
355 VFLDKFLLKNGAGNLAQIQRKRPKDGMPVSTPAQRWSTLHTEYIWDVDPQLTPPPADY 414
277 -FEDATVTVTRGSELDRVVTKHPFFYDRDSLVMNGE-----HATAEA 317
415 GTAKIHTEFP-----NMGVVTYGA---GLPNTQTNTFVSPKSG 448
318 GTGANHTAPGHGEDDPLIGKYYDLEILAPLDDRGVFTTEAPGPGFVDTANKVTEKLE 377
449 KLGGRAYVD-IVHFQPSYIDWNRSPNPGHEHPDQNSFTFAPNQGVFVS-BALYGPKL 506
378 EVGALLKMEFITHSYPHD---WRTKTP-----VIFRATAQMFASIDAFRDLLAA 424
507 LNNVLPFAPSPSQCNKPEWGQCEQAQWLKWTGEEVGDAGEIITASQHGEMVFSVGEA 566
425 VKGV-----NWT-----PAMGE-----TRLNMYDRORDGM 449
567 VSAYSSAMRLKSVYRALLNLSOTLLVVDHIEROE---DSPINSVSAPF--HNLDIDIFK 620
450 VISQRANGVP-----LPFVAENGELITDETINHISELFREHGSNVWFE 495
621 -----YIPYKPMN--RYNGAM---MDVMDAHYKMFWDHNGNSPMASI-----QEAQ 663
496 RDVKDLLPAGFTHPGSPNGEFTKETDMDV-----WFD-SGSSHQAVLNARPELSRPA 549
664 AAEFKKWTQFVNTFQWESITRIAYVFGYPIVNSCRF-IDSSNPGLOISLVNNT- 721
550 YMEGSDQYRGWEN-----SSLTTAVAITGEAPYRNVLSHGFDLGE--GRKMSKSLG 602
722 -----EHVVSIVTD-YHNKTRFNYPYVFGFGFASVA 750
603 LPKVIKVLQGADIVRLVWASVDYQADVRSDEILKQVSEYVRKXINTWRF- 653
751 DQGITRFGLGTOAI-----VKPVRHDIRIIFPPGFKFNIAVGLILC 791
654 -LGNINDFNPTNGSVYENLRVDKMYMLIKLNDLVKNVKDSYEAPEFSTIVH-QINN 711
792 ISLVILTFQWRPYLSFKLMRWILIL-----VTALWFIELLDVWSTCSQICAKWTRTE 845
712 VEL------QFYMDFAKDVVYIEADSHDRAMQTVFYEAIVTLTKLAPILPHTT--- 762
846 AEGSKKSLSSEG-----HMDLPDVVI 867
763 -EEVWNSLIGEGAESIHLDLPDVKV 787

RESULT 15
C89874
autolysin [imported] - Staphylococcus aureus (strain N315)
C;Species: Staphylococcus aureus
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C;Accession: C89874
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguchi, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; ma, A.; Hiratsuka, N.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A;Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A;Reference number: A89758; MUID:21311952; PMID:11418146

| | | | | | | | | | |
|--|-----|---|-----|--|--|--|--|--|--|
| A;Accession: C89874 | | | | | | | | | |
| A;Status: preliminary | | | | | | | | | |
| A;Molecule type: DNA | | | | | | | | | |
| A;Residues: 1-1248 <KUR> | | | | | | | | | |
| A;Cross-references: UNIPROT:Q99V41; GB:BA000018; PID:g13700854; PIDN:BAB42150.1; GSPDB:G | | | | | | | | | |
| A;Experimental source: strain N315 | | | | | | | | | |
| C;Genetics: | | | | | | | | | |
| A;Gene: atl | | | | | | | | | |
| Query Match | | | | | | | | | |
| Best local Similarity 18.3%; Score 110.5; DB 2; Length 1248; | | | | | | | | | |
| Matches 245; Conservative 146; Mismatches 462; Indels 487; Gaps 66; | | | | | | | | | |
| Qy | 30 | STPEVSNSYSEWAFVDDTDQKTO--KVDPFRPNQKLKSMLEH--SLYPDA----- | 79 | | | | | | |
| Db | 187 | TTFSASQPSVAATPKTSIPKYPQVNSINDYIRKNLAKPKIEDYTSYPPKAYRN | 246 | | | | | | |
| Qy | 80 | -----GEIQMRQKRSASHLHLP-----RAIRSAYTVMLS----- | 109 | | | | | | |
| Db | 247 | GVGRPEGIVVHDTANDRSTINGEISYMKNNYQNAFVHAFVDGDRIIETAPTDLYSWVGVA | 306 | | | | | | |
| Qy | 110 | --NPTYYLPPKH-----ADFAAKWEIYGNLPLPLALYCLLPEDKVAPE | 153 | | | | | | |
| Db | 307 | VGNPRFINVEIVHVDYASPARSMNNYADYAATQLQYYG-----LKPDS----- | 350 | | | | | | |
| Qy | 154 | FVLEYMDRMVGYKDWLVENAPG--DEV-PIGHSLTGFAFAFDFLYNLDDNHRROKYLEKI | 210 | | | | | | |
| Db | 351 | --AEYDNGVTWTHYAVSKYLGGTDHADPHGY--LRSHNYSYDQLYDLIN-----EKYLIRM | 403 | | | | | | |
| Qy | 211 | WVITEEMYEYSKVRSMGKQLLNH-----QATNMIAL-----TGAL | 247 | | | | | | |
| Db | 404 | -----GKVPWGTQTTPPTTPSKPTTSKSTGKLTVAANNVGAQIKPTNSGLY | 453 | | | | | | |
| Qy | 248 | VTGVDKGSKANIWKQAVDVNWEKTMFLNHI VDGSLDEGVAGSYTAKSVTQVVFIAQRH | 307 | | | | | | |
| Db | 454 | TTVYDKTGKA-----TNEVQKTFAV-----SKTATLGNOKFYLVQDY | 490 | | | | | | |
| Qy | 308 | FNINLNDNLKMHFWFYATLLP-----GFQRTV--GIADSNYN | 345 | | | | | | |
| Db | 491 | NSGNKF--GWKGDVVYNTAKSPVNVNQSYTSKSGTKLYTPWGTSKQVAGSVSGSGNQ | 548 | | | | | | |
| Qy | 346 | WFYGPESQLVFLDKFIL-----KNGAGNWL--AQIIRKRPKDGPM--VPSTAQRWSTLHTE | 398 | | | | | | |
| Db | 549 | TFRASKQKQI--DKSIYLYGVSNGKSGVSKAVLVDTAKEPTPIPKPSTPTTNKLTWS | 606 | | | | | | |
| Qy | 399 | YIYVDPOLTPQPADYGTAKIHFPN--WGVVYTGAGLPNTQT--NTFVSPKSGKLGRAV | 455 | | | | | | |
| Db | 607 | SL-----NGVAQINAKNGLFTTVYDKTKPTKEVQKTFVVTKEASLGSKNF | 653 | | | | | | |
| Qy | 456 | YDIVHFQPSWIDGWRSFNPGHEHPDONSFTFAPNGQVFVSEALYGPKLSHLNNLVFAP | 515 | | | | | | |
| Db | 654 | YLV-----KDYNSPTLI--GWVKQGDVIYNNAKSPVNMQTYTV | 690 | | | | | | |
| Qy | 516 | SPSSQC--NKPWEGQGECAQWLKWTGEEVGDAAAGEIITASQHGEMFV-----SCEAV | 567 | | | | | | |
| Db | 691 | KPGTKLYSVFW--GTYKQEGAVSGTGNQTFKATKQ-----QQIDKSIYLFGTVNGKSGWS | 745 | | | | | | |
| Qy | 568 | SAY-----SSAMRLKSVYRALLLNSQTLIVVDHIERQEDSPINSVSAFFHNLDIDP | 619 | | | | | | |
| Db | 746 | KAVLAVPAAPKKAQAQPKTAVKATVYTKPTQTOTVSKIAQVK--PNNT----- | 791 | | | | | | |
| Qy | 620 | KYIPYKEMRYNGAMMDVDADHYKMFDFHGHGNSPMASIQEASQAAEFKKRWTFQFVNVT | 679 | | | | | | |
| Db | 792 | -----GIRASVYE-----KTAKNGAKYADR-----TF | 813 | | | | | | |
| Qy | 680 | QMBSTITRIAYFYGPYINVSSCRFIDSNNPGLQISLVNNTBHVSVISVTDYHNLKTRFN | 739 | | | | | | |
| Db | 814 | -----YVTKERAH-----GNETYVLLNNTSH--NIPLGWENVKD-LN | 847 | | | | | | |
| Qy | 740 | YLGFGGFSVADQGIITRFGLTQAIIVKPVVRHRIIPFGFK-----FNIAVGLILCI | 792 | | | | | | |
| Db | 848 | VQNLGKEVKTKTKYTVNKSNNGLSMV-----PWGTKNQVILITGNNTIAQGTFNAT | 896 | | | | | | |

Search completed: June 23, 2005, 08:50:24
Job time : 34.7171 secs

| | | | | | | | | | |
|----|------|--|------|--|--|--|--|--|--|
| Qy | 793 | SLVILTQWRPYLSPFKLMRWILILVIALWFIELDVMSTCSQPICAKWRTRTEAGSKS | 852 | | | | | | |
| Db | 897 | KQVSVGKDVLYLTINNRTGWNA-----KDLTAPTAVKPTTSAAKDYNVT | 942 | | | | | | |
| Qy | 853 | L---SSEGHMDLP--DVVITSLPGSGAEILKQLFNSSDFLYIRVPTAYIDIPETELEI | 907 | | | | | | |
| Db | 943 | YVIKNGNGYVYVTPNSDTAKYSLKA-----FNEQPFVAVK----- | 977 | | | | | | |
| Qy | 908 | DSFVDACEWKVSDIRSGHFRLLRGWLSLQVDTKLHLQNIHLHPEPNRGKLAQYFAMKDK | 967 | | | | | | |
| Db | 978 | EQVINGQWYKLSNGKU-----AWIKS-----TDLAKELI----- | 1009 | | | | | | |
| Qy | 968 | KKFKRRRESLPEQRSMQMGAFDRDAEYIRALRRHLVYYPSPARPVLSLSSGSWTLKLFHQ | 1027 | | | | | | |
| Db | 1010 | --KYNQGTMTLNQVAQIQAG-----LQYKPVQVRV-----PGKWTDA--NFN | 1047 | | | | | | |
| Qy | 1028 | EVLGASMRALYIVRDPRAWIYMSLYNSKPSLYSLKNVPBHLAKLPKIEGGKCKML--NSG | 1086 | | | | | | |
| Db | 1048 | DVKHA-MDTKRLAQDP-ALKYQFLRLDQPNISIDKINQFL-----KGGVLENQG | 1096 | | | | | | |
| Qy | 1087 | YAFVEYELRKELSKSNAVSLSLHLWLANTAAALRINTDLLPT---SYQLVKFEDIVHF | 1143 | | | | | | |
| Db | 1097 | AAF-----NKAQMYGINEVYLISHA-----LLETNGTSQLAGADVVN- | 1136 | | | | | | |
| Qy | 1144 | PQKTTERTIFAFILGPLSPASLNQILFATSTNLFYLPYEGEISPTNTNWKQNLPRDEIKL | 1203 | | | | | | |
| Db | 1137 | -----NKVV--TNSNTKYHNVFGLAAAYDNDPL-----REGIKY | 1167 | | | | | | |
| Qy | 1204 | IENICWTLMDR--LGYPKFM | 1221 | | | | | | |
| Db | 1168 | AKQAGWDTVSKAIVGGAKEFI | 1187 | | | | | | |

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2005, 08:33:13 ; Search time 134.102 Seconds
(without alignments)
4666.302 Million cell updates/sec

Title: US-10-697-828-15

Perfect score: 6532

Sequence: 1 MPKGGAPPWIMALMFTGHL.....LIENICWLMRLGYPKFM 1222

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

UniProt_03:*

1: uniprot_sprot:*

2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|--------------|---------------------|
| 1 | 6527 | 99.9 | 1222 | 2 Q8IZU8 | Q8IZU8 homo sapien |
| 2 | 4186 | 64.1 | 784 | 2 Q5P523 | Q5P523 homo sapien |
| 3 | 4037 | 61.8 | 755 | 2 Q3HAK5 | Q3HAK5 homo sapien |
| 4 | 2983.5 | 45.7 | 597 | 2 Q8BZP3 | Q8BZP3 mus musculus |
| 5 | 2597 | 39.8 | 480 | 2 Q9H951 | Q9H951 homo sapien |
| 6 | 1757.5 | 26.9 | 958 | 2 Q8BL14 | Q8BL14 m mus muscu |
| 7 | 1753.5 | 26.8 | 958 | 2 Q9UL01 | Q9UL01 homo sapien |
| 8 | 1331 | 20.4 | 314 | 2 Q9HAD7 | Q9HAD7 homo sapien |
| 9 | 1242 | 19.0 | 230 | 2 Q6J9G8 | Q6J9G8 gorilla gor |
| 10 | 1194 | 18.3 | 222 | 2 Q6J9G9 | Q6J9G9 pan troglod |
| 11 | 251.5 | 3.9 | 680 | 2 Q8CJW4 | Q8CJW4 streptomyce |
| 12 | 192 | 2.9 | 776 | 2 Q8UBJ1 | Q8UBJ1 agrobacteri |
| 13 | 150 | 2.3 | 411 | 2 Q9EQC0 | Q9EQC0 mus musculu |
| 14 | 147 | 2.3 | 411 | 2 Q43916 | Q43916 homo sapien |
| 15 | 145 | 2.2 | 2504 | 2 Q85160 | Q85160 photorhabdu |
| 16 | 143.5 | 2.2 | 761 | 2 Q9KWR4 | Q9KWR4 spingomona |
| 17 | 134 | 2.1 | 483 | 2 Q794G9 | Q794G9 mus musculu |
| 18 | 134 | 2.1 | 530 | 2 Q88276 | Q88276 mus musculu |
| 19 | 134 | 2.1 | 530 | 2 Q80WV3 | Q80WV3 mus musculu |
| 20 | 133 | 2.0 | 483 | 2 Q9UED5 | Q9UED5 homo sapien |
| 21 | 133 | 2.0 | 530 | 2 Q9Y4C5 | Q9Y4C5 homo sapien |
| 22 | 132.5 | 2.0 | 1337 | 2 Q8A3U4 | Q8A3U4 bacteroides |
| 23 | 129 | 2.0 | 304 | 2 Q677T4 | Q677T4 lymphocyeti |
| 24 | 128.5 | 2.0 | 532 | 2 Q644R5 | Q644R5 erwinia car |
| 25 | 128 | 2.0 | 486 | 2 Q9VMC3 | Q9VMC3 drosophila |
| 26 | 127 | 1.9 | 441 | 2 Q93403 | Q93403 torpedo cal |
| 27 | 127 | 1.9 | 1456 | 1 MANR_HUMAN | P22897 homo sapien |
| 28 | 127 | 1.9 | 2958 | 2 Q7RNZ3 | Q7RNZ3 plasmodium |
| 29 | 126.5 | 1.9 | 420 | 2 Q6DBY9 | Q6DBY9 brachydanio |
| 30 | 126 | 1.9 | 388 | 2 Q9WUE5 | Q9WUE5 mus musculu |
| 31 | 126 | 1.9 | 2019 | 2 Q6BEW0 | Q6BEW0 caenorhabdi |

ALIGNMENTS

RESULT 1

| ID | Q8IZU8 | PRELIMINARY; | PRT; | 1222 AA. |
|----|---|--------------|------|----------|
| AC | Q8IZU8; | | | |
| DT | 01-MAR-2003 (Tremblrel. 23, Created) | | | |
| DT | 01-MAR-2003 (Tremblrel. 23, Last sequence update) | | | |
| DT | 01-MAR-2004 (Tremblrel. 26, Last annotation update) | | | |
| DE | NCAG1. | | | |
| OS | Homo sapiens (Human). | | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. | | | |
| OX | NCBI_TaxID=9606; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RA | Goossens D., Del-Pavero J., Van Broeckhoven C.; | | | |
| RL | Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases. | | | |
| DR | EMBL; AF480435; AAN32895.1; - | | | |
| DR | GO; GO:0008146; F:sulfotransferase activity; IEA. | | | |
| DR | InterPro; IPR000863; Sulfotransferase. | | | |
| DR | Pfam; PF00685; Sulfotransfer 1; 1. | | | |
| SQ | SEQUENCE 1222 AA; 140272 MW; 156BE397B4E99AE9 CRC64; | | | |

| | | | | | | | |
|-----------------------|--------|--|-------|------------|----|--------|-------|
| Query Match | 99.9%; | Score | 6527; | DB | 2; | Length | 1222; |
| Best Local Similarity | 99.9%; | Pred. No. | 0; | | | | |
| Matches | 1221; | Conservative | 0; | Mismatches | 1; | Indels | 0; |
| Gaps | 0; | | | | | | |
| Qy | 1 | MPKGGAPPWIMALMFTGHLFLALLMFAFTFEESVSNYSEWAVFTDDIDQFKTKVQDF | 60 | | | | |
| Db | 1 | MPKGGAPPWIMALMFTGHLFLALLMFAFTFEESVSNYSEWAVFTDDIDQFKTKVQDF | 60 | | | | |
| Qy | 61 | RPNQKLKSMHPSLYPDAGEIQAMRQKSRASHLHLFRAIRSAVTMLSNPTLYLPPPKH | 120 | | | | |
| Db | 61 | RPNQKLKSMHPSLYPDAGEIQAMRQKSRASHLHLFRAIRSAVTMLSNPTLYLPPPKH | 120 | | | | |
| Qy | 121 | ADFAAKWNEIYGNLPPALYCLLPEDKVAFEVLEIMDMVGYKDWLVENAFGEVPI | 180 | | | | |
| Db | 121 | ADFAAKWNEIYGNLPPALYCLLPEDKVAFEVLEIMDMVGYKDWLVENAFGEVPI | 180 | | | | |
| Qy | 181 | GHSITGTFATFDLYNLNDNRKQKLEKIWIITEEMVEYSKVRSGKQLLNHQATNMI | 240 | | | | |
| Db | 181 | GHSITGTFATFDLYNLNDNRKQKLEKIWIITEEMVEYSKVRSGKQLLNHQATNMI | 240 | | | | |
| Qy | 241 | ALLTGLVTVGDKGSKANIWKQAVDVMEKTMFLNHNHIVDGLDEGVAYGSYTAKSVTQY | 300 | | | | |
| Db | 241 | ALLTGLVTVGDKGSKANIWKQAVDVMEKTMFLNHNHIVDGLDEGVAYGSYTAKSVTQY | 300 | | | | |
| Qy | 301 | VFLAQRHFNINLNDNNLKMHFYFYATLLPGPQRTVGIADSNYNWVFGPSQLVFLDKF | 360 | | | | |
| Db | 301 | VFLAQRHFNINLNDNNLKMHFYFYATLLPGPQRTVGIADSNYNWVFGPSQLVFLDKF | 360 | | | | |
| Qy | 361 | ILKNAGNWLAAQIRKRPKDGMPVSTAQRWSTLHTEYIWDLPQTPQPPADYGTAKIH | 420 | | | | |
| Db | 361 | ILKNAGNWLAAQIRKRPKDGMPVSTAQRWSTLHTEYIWDLPQTPQPPADYGTAKIH | 420 | | | | |

RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Diencephalon;
RA The FANTOM Consortium
RT the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs";
RL Nature 420:563-573(2002).
[4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Diencephalon;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.,
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT Prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
[5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Diencephalon;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsunoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa K., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.,
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
[6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Diencephalon;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ono M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shingawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.,
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK033983; BAC28538.1; -
DR MGD; MGI:2424948; 9330132B09Rik.
DR GO; GO:0005615; C:extracellular space; TAS.
FT NON TER 597 597
SQ SEQUENCE 597 AA; 68197 MW; E344P433EB16EDDC CRC64;
Query Match 45.7%; Score 2983.5; DB 2; Length 597;
Best Local Similarity 90.0%; Pred. No. 6.6e-208;
Matches 542; Conservative 32; Mismatches 23; Indels 5; Gaps 1;
QY 11 MALMFTCHLLFLALLMFASTFEESVSNSEYSAVFTDDIDQFTQKQVDPNPQKLKSM 70
DB 1 MAFMFTBHLFLTLMMCSFTCESVSNSEYSAVFTDDIQWLKSKIQDFLNR----- 55
QY 71 LHPSLYFDAGEIQAMRQKSRASHLHLFRAIRSAVTVMLSNPTLYLPPPKHADFAAKWNEI 130
DB 56 LHPNLVFDAGDIQTLQKSRTHLHIFRAIKSAVTIMLSNPSYLLPPPKHAEFAAKWNEI 115
QY 131 YGNLPLALYCLLCPDKVAFVFLVMDRMVGYKDWLVENAPGVDPVIGHSLTGATA 190
DB 116 YGNLPLALYCLLCPDKVAFVFLVMDRMVGYKDWLVENAPGVDPVIGHSLTGATA 175
QY 191 FDFLYNLLDNRHQRKYLEKIWIWTEEMVEYSKVRSGKQLLHQNATNMIALLTGALVTG 250

DB 176 FDFLYNLLGQRKQKYLEKIWIWTEEMVEYSKIRSGKQLLHQNATNMIALLTGALVTG 235
QY 251 VDKGSKANIWKQAVVDVMEKTMFLNLHIVDGLSDGVAYGYSYAKSVTQYVFLAQRHFN 310
DB 236 VDKGSKANIWKQAVVDVMEKTMFLNLHIVDGLSDGVAYGYSYAKSVTQYVFLAQRHFN 295
QY 311 NNLDNNLWKQHFYFYATLLPGFQRTVGTADSNYNWFYGPESQLVFLDKFLLKNGAGNWL 370
DB 296 NNFDNNLWKQHFYFYATLLPGYQRTVGLADSNYNWFYGPESQLVFLDKFLLQNGAGNWL 355
QY 371 AQQIRKRPKQGMVPESTQAQRMSTLHTEYIWDYDQLTQPPADYGTAKIHTHPNMGVVTY 430
DB 356 AQQIRKRPKQGMVPESTQAQRMSTLHTEYIWDYDQLTQPPVDFGTAKMHTFPNMGVVTY 415
QY 431 GAGLPNTQNTVPSFKSGKLGGRVYDVHFPYPSWIDGWRSNFPGCHEHPDQNSFTFAPN 490
DB 416 GCGLPNTQNTVPSFKSGKLGGRVYDVHFPYPSWIDGWRSNFPGCHEHPDQNSFTFAPN 475
QY 491 GQFVSEALYGPKLHNLNVLPAPSPSOCKNPHGOLGECANLKWTEGVGDAAGEI 550
DB 476 GQFVSEALYGPKLHNLNVLPAPSPSOCKNPHGOLGECANLKWTEGVGDAAGEV 535
QY 551 ITASQHGEMVFSGEAVSAYSSAMELKSIVYRALLNLSQTLVVDHIERQEDSPINSVA 610
DB 536 ITAQHGDMFVSGEAVSAYSSAMELKSIVYRALLNLSQTLVVDHIERQETSPINSVA 595
QY 611 FF 612
DB 596 FF 597
RESULT 5
Q9H951 PRELIMINARY; PRT; 480 AA.
AC Q9H951;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ13005.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
RX PubMed=14702039; DOI=10.1038/ng1285;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahari K.,
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA Sudo H., Hosiiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,
RA Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,
RA Tanai H., Kimata M., Watanabe M., Hiraoaka S., Chiba Y., Ishida S.,
RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,
RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,
RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,
RA Musashino K., Yuuki H., Oshima A., Sasaki N., Aoteuka S.,
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wake H.,
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,

RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
RT Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.,
RT "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs";
RL Nat. Genet. 36:40-45(2004).
DR EMBL; AK023067; BAB14387.1; -;
SQ SEQUENCE 480 AA; 54559 MW; AEF246851B83E9DB CRC64;

Query Match 39.8%; Score 2597; DB 2; Length 480;
Best Local Similarity 99.8%; Pred. No. 5.9e-180;
Matches 479; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 320 MHFWFYATLLPGFQRTVGIADSNYNFYGPESQVLFDLKFILKXGAGNWLAAQIRKHP 379
Db 1 MHFWFYATLLPGFQRTVGIADSNYNFYGPESQVLFDLKFILKXGAGNWLAAQIRKHP 60
Qy 380 KDGPMVSTAQRWSTLHTEYIWDYDQLTTPQPPADYGTAKIHTFFNWGVVYTGAGLPNTQT 439
Db 61 KDGPMVSTAQRWSTLHTEYIWDYDQLTTPQPPADYGTAKIHTFFNWGVVYTGAGLPNTQT 120
Qy 440 NTFVSPKSGKLGRAVVDIVHFQYPMIDGWRSPNGHEHPDQNSFTFAPNGQVVFSEAL 499
Db 121 NTFVSPKSGKLGRAVVDIVHFQYPMIDGWRSPNGHEHPDQNSFTFAPNGQVVFSEAL 180
Qy 500 YGPKLSHLNNVLPAPSPSSQCNKPWEGQGLGECAQWLKWTGEYVGDAAGEIITASQHGEM 559
Db 181 YGPKLSHLNNVLPAPSPSSQCNKPWEGQGLGECAQWLKWTGEYVGDAAGEIITASQHGEM 240
Qy 560 VFVSGEAVSAYSSAMRLKSVYRALLLNLSQTLVVDHIERQEDSPINSVSAPFNLDIDF 619
Db 241 VFVSGEAVSAYSSAMRLKSVYRALLLNLSQTLVVDHIERQEDSPINSVSAPFNLDIDF 300
Qy 620 KYIPYKFNRRNGAMVDVMDAHYKMFDFHNGSPMASIQAEQAQAEFKKRWTOFVNVT 679
Db 301 KYIPYKFNRRNGAMVDVMDAHYKMFDFHNGSPMASIQAEQAQAEFKKRWTOFVNVT 360
Qy 680 QMESTITRIAYVFYGPYINVSCHRFIDSSNPGQLQSLNNVNTHEVSVITVDYHNLKTRFN 739
Db 361 QMESTITRIAYVFYGPYINVSCHRFIDSSNPGQLQSLNNVNTHEVSVITVDYHNLKTRFN 420
Qy 740 YLFGGFGFASVADQGGITRFLGTOAIVKPVVRHRIIPFPFGKFNIAVGLILCISLVILTF 799
Db 421 YLFGGFGFASVADQGGITRFLGTOAIVKPVVRHRIIPFPFGKFNIAVGLILCISLVILTF 480

RESULT 6
Q8BL14 PRELIMINARY; PRT; 958 AA.
AC Q8BL14
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 25-OCT-2004 (TREMBLrel. 28, Last annotation update)
DE Mus musculus 9.5 days embryo parthenogenote cDNA, RIKEN full-length
DE enriched library, clone: B130024B19 product: SQUAMOUS CELL CARCINOMA
DE ANTIGEN RECOGNIZED BY T CELL homolog (Squamous cell carcinoma antigen
DE recognized by T cells 2).
GN Name: Sert2;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Parthenogenote;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Parthenogenote;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;

RT "Functional annotation of a full-length mouse cDNA collection";
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Parthenogenote;
RX The FANTOM Consortium;
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs";
RL Nature 420:563-573(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Parthenogenote;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Parthenogenote;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Parthenogenote;
RX Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Okamoto N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toyota T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,
RA Klausner R.D., Collins J.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalish D.E., Schnerker A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [8]


```
Db 342 LADQIRNRVVGEGTSGKQRMCTLHTEFLWYDGLSKSVPPDPFGTPTLHFFDGMVVT 401
Qy 430 YGAGLPNTQNTFVSFKSGKLGRAVDYIVHFQPY-SWIDGWRSPNPGHHPDQNSFTEA 488
Db 402 YGALPAEINRSFLSKSGKLGGRAYIDIVHRNKKYKDWIKWRNFGHAGHEHPDQNSFTEA 461
Qy 489 PNGQVVFSEALYGPGLSHLNNLVFAPSPSCQNKPMWQGLGE--CAQWLKWTGEEYVGD 546
Db 462 PNGVVFTEALYGPKYTFNNVLMFSPAVSKSCFSPWVGQVTEDCSSKWSKYKHDLAASC 521
Qy 547 AGEIITASQHGEMVFSGEAVSAVSSAMRLKSVYRALLLNSQTLVVDHIEROEDSPIN 606
Db 522 QGRVVAEEKNGVVFIRGEGVGAYNPQLNKNVORNILLHPQLLLVDQHLGEESPLE 581
Qy 607 SVSAFFENLDDIFKVIPIYKFMNRYNGAMMDVDAHYKMFWDHGHGNSPMASIQEAAAE 666
Db 582 TAASFFHNVDVFPF---ETVVDGVGHGAFIRQRDGLYKNYMMDDTGYSEKATFASVTPRG 638
Qy 667 FKRWTFQVNVTFQWESTITRIAYVFGPYINVSSCRFDIDSSNPGLOQISLNNVNTHEVWS 726
Db 639 YPYNGTNYVNVTHLRSPITRAAYLFGPSIDVQS-----FTVHGDSQQLDVF 686
Qy 727 IVTDYHNLKTRFNLYL-----GFGGFASV-ADQSQI 755
Db 687 IATSKHAYAT---YLWTGEATGSAFAQVIADRHKI 719

RESULT 8
Q9HAD7
ID AC Q9HAD7 PRELIMINARY; PRT; 314 AA.
AC Q9HAD7
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE Hypothetical protein FLJ11787.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSU=whole embryo;
RX PubMed=14702039; DOI=10.1038/ng1285;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Ohtsuka M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA Sudo H., Hosiiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,
RA Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,
RA Tanai H., Kinata M., Watanabe M., Hiraoa K., Chiba Y., Ishida S.,
RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,
RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,
RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,
RA Yashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
RA Moriya S., Momiya H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Ohtani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsumura K., Nakajima Y., Munzo T., Morinaga M., Sasaki M.,
RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isozaki T., Sugano S.;
RA "Complete sequencing and characterization of 21,243 full-length human
```

```
RT cDNAs.";
RL Nat. Genet. 36:40-45 (2004).
DR EMBL; AK021849; BAB13912.1; -.
FT NON TER 314 314
SQ SEQUENCE 314 AA; 35955 MW; 722B5555DC065538 CRC64;

Query Match
Best Local Similarity 32.4%; Score 1331; DB 2; Length 314;
Matches 313; Conservative 0; Mismatches 1; Indels 652; Gaps 1;

Qy 1 MPKGGAPPMTALMFTGHLFLALLMPARFTEESVSNYSEWAVFTDDIDQFKTKQVQDF 60
Db 1 MPKGGAPPMTALMFTGHLFLALLMPARFTEESVSNYSEWAVFTDDIDQFKTKQVQDF 60
Qy 61 RPNQKLKSKMLHPSLYFDAGEIQAMRQKSRASHLHLFRAIRSAVTVMLSNPTYYLPPPKH 120
Db 61 RPNQKLKSKMLHPSLYFDAGEIQAMRQKSRASHLHLFRAIRSAVTVMLSNPTYYLPPPKH 120
Qy 121 ADFAAKWEIYGNLPLALYCLLPEDKVAPEFVLYMDRMVGYKDWLVENAPGDEVPI 180
Db 121 ADFAAKWEIYGNLPLALYCLLPEDKVAPEFVLYMDRMVGYKDWLVENAPGDEVPI 180
Qy 181 GHSITGFATAFDFLYNLLDNHRRQKYLEKIWIITEEMYEYSKVRSGKQLLHNNHOATMI 240
Db 181 GHSITGFATAFDFLYN----- 196
Qy 241 ALLTGALVTGVDKSGKANIWKQAVVDVMEKTMFLNHLNIVDGLDEGVAYGYSYTAQSVTQY 300
Db 197 ----- 196
Qy 301 VFLAQRHFNINLNDNNLWKHFMYFYATLLPGFQRTVGIADSNYNWYGPESQLVFLDKF 360
Db 197 ----- 196
Qy 361 ILKNGAGWLAQIRKRPKDGMPVSTAQRWSTLHTEYIWDYDQLTPQPPADYGTAKIH 420
Db 197 ----- 196
Qy 421 TFPNWGVVTVYAGLPNTQNTFVSFKSGKLGRAVDYIVHFQPYSWIDGWRSPNPGHEHP 480
Db 197 ----- 196
Qy 481 DQNSFTFAPNGQVVFSEALYGPGLSHLNNLVFAPSPSCQNKPMWQGLGECAQWLKWTG 540
Db 197 ----- 196
Qy 541 BEVGDAAGEIITASQHGEMVFSGEAVSAVSSAMRLKSVYRALLLNSQTLVVDHIERQ 600
Db 197 ----- 196
Qy 601 EDSPIINSVSAFFHNLDDIFKVIPIYKFMNRYNGAMMDVDAHYKMFWDHGHGNSPMASIQE 660
Db 197 ----- 196
Qy 661 AEOAAEFKRWTFQVNVTFQWESTITRIAYVFGPYINVSSCRFDIDSSNPGLOQISLNNV 720
Db 197 ----- 196
Qy 721 TEHVSVIVTDYHNLKTRFNLYGFGGFASVADQSQITRFGLTQAIKVPVRHDIRIIFPGP 780
Db 197 ----- 196
Qy 781 KFNIAVGLILCISILVILTFQWRVFLSPKLMRMWILILVIALWFLTELLDVMWSTCSQICAK 840
Db 197 ----- 196
Qy 841 WTRTEAGSKSLSSSEGHMDLPDVVITSPLPGSGAEILKOLFFNSSDFLYIRVPTAYIDI 900
Db 197 -----CKKSLSSSEGHMDLPDVVITSPLPGSGAEILKOLFFNSSDFLYIRVPTAYIDI 248
Qy 901 PETELEIDSFVDAACEWKVSDIRSGHGFRLRLGWLQSLVQDTKLHLQNLHLPENRGKLAQY 960
Db 249 PETELEIDSFVDAACEWKVSDIRSGHGFRLRLGWLQSLVQDTKLHLQNLHLPENRGKLAQY 308
```

```

QY 961 FAMNKD 966
DB 309 FAMNKD 314

RESULT 9
Q6J9G8 PRELIMINARY; PRT; 230 AA.
AC Q6J9G8; 2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE NCAG1 (Fragment).
OS Gorilla gorilla (gorilla).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Gorilla.
OX NCBI_TaxID=9593;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=15123584; DOI=10.1101/gr.1891104;
RA Zhang J., Wang X., Podlaha O.;
RT "Testing the chromosomal speciation hypothesis for humans and
chimpanzees.";
RL Genome Res. 14:845-851(2004).
DR EMBL: AY561496; AAT45534.1; -.
FT NON_TER 230
FT NON_TER 230
SQ SEQUENCE 230 AA; 26793 MW; 1387502DB633B3A3 CRC64;

Query Match 19.0%; Score 1242; DB 2; Length 230;
Best Local Similarity 100.0%; Pred. No. 6.3e-82;
Matches 230; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 FEFVLEMDRMVGYKDWLVENAPGDEVPIGHSLTGFATFDFLYNLLDNHRRQKYLEKI 211
DB 1 FEFVLEMDRMVGYKDWLVENAPGDEVPIGHSLTGFATFDFLYNLLDNHRRQKYLEKI 60

QY 212 VITEEMEYKSVRSWKGKOLLHNHQTNNMIALLTGALVTGVDKSGKANIWKQAVVDVMEKT 271
DB 61 VITEEMEYKSVRSWKGKOLLHNHQTNNMIALLTGALVTGVDKSGKANIWKQAVVDVMEKT 120

QY 272 MFLNLHIVDGLSDEGVAYGYSYTAKSVTQYVFLAQRHFNNLNNDNNLWKHFYFYATLLP 331
DB 121 MFLNLHIVDGLSDEGVAYGYSYTAKSVTQYVFLAQRHFNNLNNDNNLWKHFYFYATLLP 180

QY 332 GFQRTVGIADSNYNWFYGPESQLVFLDKFILKNGAGNWLAAQIIRKRPXD 381
DB 181 GFQRTVGIADSNYNWFYGPESQLVFLDKFILKNGAGNWLAAQIIRKRPXD 230

RESULT 10
Q6J9G9 PRELIMINARY; PRT; 222 AA.
AC Q6J9G9; 2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE NCAG1 (Fragment).
OS Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX NCBI_TaxID=9598;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=15123584; DOI=10.1101/gr.1891104;
RA Zhang J., Wang X., Podlaha O.;
RT "Testing the chromosomal speciation hypothesis for humans and
chimpanzees.";
RL Genome Res. 14:845-851(2004).
DR EMBL: AY561495; AAT45533.1; -.
FT NON_TER 222
FT NON_TER 222

SQ SEQUENCE 222 AA; 25736 MW; 4C0CCB8E31AF6CA8 CRC64;

Query Match 18.3%; Score 1194; DB 2; Length 222;
Best Local Similarity 100.0%; Pred. No. 1.9e-78;
Matches 222; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 155 VLEYMDRMVGYKDWLVENAPGDEVPIGHSLTGFATFDFLYNLLDNHRRQKYLEKI 214
DB 1 VLEYMDRMVGYKDWLVENAPGDEVPIGHSLTGFATFDFLYNLLDNHRRQKYLEKI 60

QY 215 EMEYKSVRSWKGKOLLHNHQTNNMIALLTGALVTGVDKSGKANIWKQAVVDVMEKT 274
DB 61 EMEYKSVRSWKGKOLLHNHQTNNMIALLTGALVTGVDKSGKANIWKQAVVDVMEKT 120

QY 275 LNHIIVDGLSDEGVAYGYSYTAKSVTQYVFLAQRHFNNLNNDNNLWKHFYFYATLLP 334
DB 121 LNHIIVDGLSDEGVAYGYSYTAKSVTQYVFLAQRHFNNLNNDNNLWKHFYFYATLLP 180

QY 335 RTVGIADSNYNWFYGPESQLVFLDKFILKNGAGNWLAAQIIRK 376
DB 181 RTVGIADSNYNWFYGPESQLVFLDKFILKNGAGNWLAAQIIRK 222

RESULT 11
Q8CJW4 PRELIMINARY; PRT; 680 AA.
AC Q8CJW4; 2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein SCO3500.
GN ORFNames=SC1134.01c, SC565.36c;
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NA3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953; DOI=10.1038/417141a;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
Huang C.-H., Kieser T., Larke L., Murphy L.D., Oliver K., O'Neil S.,
Rabinowitch E., Rajandream M.A., Rutherford K.M., Rutter S.,
Seeger K., Saunders D., Sharp S., Squares S., Taylor K.,
Warren T., Wietzorrek A., Woodward J.R., Barrell B.G., Parkhill J.,
Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL: AL939116; CAD5321.1; -.
DR PIR: T36083; T36083.
KW InterPro: IPR008929; Chondroitin lyase.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 680 AA; 75582 MW; 631B150D802EF81C CRC64;

Query Match 3.9%; Score 251.5; DB 2; Length 680;
Best Local Similarity 22.6%; Pred. No. 3.1e-09;
Matches 125; Conservative 86; Mismatches 249; Indels 93; Gaps 27;

QY 80 GEIQAMRQKSRASHLHLFRATRSATVMSLNPTYYL---PPPKHADFAKWNIEYGNL 136
DB 6 GRDLSRLDGLNGTHAAQWRRLHEQC-----DWYRRQNPPTEPEASITY---FGPAAA 55

QY 137 PLALYCLLPCKVAFEFVLEYMDRMVGYKDWLVENAPGDEVPIGHSLTGFATFDFLYN 196
DB 56 NLALAYRLTQ--RGYLEEAWRWISVCVAYPHWGRAHMDHDLDAAGLLHGLSLAWSLGE 114

QY 197 LLDNHRQKYLEKIWIITEEMEYKSV---RSWGKOLLHNHQTNNMIALLTGALVTGVDK 253
DB 115 DLEPERREILRAKLBLOGERLHSPABETTGWRWSSAYWQNH---NWIC-WTGIATAGVAL 170

```


OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
ON NCBI_TaxID=10090;
OX [1]
RY SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Mammary gland;
RX MEDLINE=21096027; PubMed=11181564; DOI=10.1093/glycob/11.1.75;
RA Hemmerich S., Lee J.-K., Bhakta S., Bistrup A., Ruddle N.R.,
RO Rosen S.D.,
RT "Chromosomal localization and genomic organization of the galactose/N-
acetylglucosamine/N-acetylglucosamine 6-O-sulfotransferase gene
family";
RL Glycobiology 11:75-87(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Mammary gland;
RX MEDLINE=22389257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalka J., Schmutz J., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Mammary gland;
RA Strausberg R.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF280087; AAC48245.1; -.
DR EMBL: BC030667; AAH30667.1; -.
DR MGD; MGI:1924219; Chst1.
DR GO; GO:0005615; C-extracellular space; TAS.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
KW Transferase.
SQ SEQUENCE 411 AA; 46903 MW; B1AB590BF5B9CBDC CRC64;
Query Match 2.3%; Score 150; DB 2; Length 411;
Best Local Similarity 18.0%; Pred. No. 0.036;
Matches 88; Conservative 56; Mismatches 154; Indels 190; Gaps 14;
QY 813 WILVIALWFIEL----LDVWSTCSOPICAKWTRTEAGSKSLSEGHMDL-----P 863
DB 5 WKAVLLALLASIAIQTAYTFTAKSFHCPLGTDT---GLAERLCEEGFTSYNLSRKT 61
QY 864 DVVITSLPGSGAILKQLFPNSSDFYIRVPTAYIDIPETELEIDSFDVACEWKVSDIRS 923
DB 62 HVLILATTRSGSSFGVGLFNQHDVFLFEPL----- 93
QY 924 GHFELLRGWLQSLVDQTLKHLQNLHLHPNKGKLAQYFAMNKDKKFKRRSLPQRQ 983
DB 94 -----YHVQNTLIPRFTQG-----KSPADRRVWLGASRDL 123
QY 984 MKGAFFDRDAEYI-----RALRRHLVYVPSARPV-----LSLSSGSWTL 1021
DB 124 LRSIYDCDLYFLENYIKPPPNVHTNRVFRGASVLCSPVCPDPSGSDLLILEGDCVR 183
QY 1022 KLHFFQEVLGA-----SMRALYIVRDPRAWISM 1050
DB 184 MCGLLNLTAAACRERSHVAIKTVRVPEVNDLRALVEDPRLNLKVLQIVRDPRGILASR 243

QY 1051 LYNKPSLSYSLKNVPEHLAKLFKIEGGKGNLNSGYAFEYEPRLKELSK-----SK 1102
DB 244 SETPRDY-----RLWRLWYGTGR-----KPYNDLVTLTTCEDFS 281
QY 1103 SNAVLSLHLWLANTAALRINTDLLPTSQYLVKPEDIVHFPQKTERIFAFGLIPL---- 1159
DB 282 SVSTGLMRPWSLKG-----KYMVLRYEDLARNPMKTEIYEFLGIPLD 327
QY 1160 -----SPASLNQILPATSTNLFLPYEGETSPNTNWKQNLPRDEIKLIENICW 1209
DB 328 VAHWIQNNTRGPTLGGHKYSTVRN-----SAATAEKWRPRLSYDIVAFQAQACQ 377
QY 1210 TLMDRIGY 1217
DB 378 QVLAQLGY 385
RESULT 14
O43916 PRELIMINARY; PRT; 411 AA.
AC O43916;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 25-OCT-2004 (TREMBLrel. 28, Last annotation update)
DE Chondroitin-6-sulfotransferase (Keratan sulfate Gal-6-
sulfotransferase) (Carbohydrate sulfotransferase 1) (Carbohydrate
(Keratan sulfate Gal-6) sulfotransferase 1).
GN Name=CHST1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98306085; PubMed=9639683; DOI=10.1016/S0925-4439(98)00028-3;
RA Mazany K.D., Peng T., Watson C.E., Tabas I., Williams K.J.;
RT "Human chondroitin 6-sulfotransferase: cloning, gene structure, and
chromosomal localization.";
RL Biochim. Biophys. Acta 1407:92-97(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97189336; PubMed=9117134;
RA Williams K.J.;
RT "Atherosclerosis: cell biology and lipoproteins.";
RL Curr. Opin. Lipidol. 7:U202-U208(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98070405; PubMed=9405439; DOI=10.1074/jbc.272.51.32321;
RA Fukuta M., Inazawa J., Torii T., Tsuzuki K., Shimada E., Habuchi O.;
RT "Molecular cloning and characterization of human keratan sulfate Gal-
6-sulfotransferase.";
RL J. Biol. Chem. 272:32321-32328(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Umbilical vein endothelium;
RX MEDLINE=99168906; PubMed=10045991; DOI=10.1006/geno.1998.5653;
RA Li X., Tedder T.F.;
RT "CHST1 and CHST2 sulfotransferases expressed by human vascular
endothelial cells: cDNA cloning, expression, and chromosomal
localization.";
RL Genomics 55:345-347(1999).
RN [5]
RP SEQUENCE FROM N.A.
RA Rieder M.J., Livingston R.J., Daniels M.R., Chung M.-W.,
RA Miyamoto K.E., Nguyen C.P., Nguyen D.A., Poel C.L., Robertson P.D.,
RA Schackwitz W.S., Sherwood J.K., Witrak L.A., Nickerson D.A.;
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;

RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahney J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalhus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Maiz M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [7]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strauberg R.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
RN [8]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strauberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; U65637; AAC28776.1; -;
DR EMBL; AB003791; BAA24840.1; -;
DR EMBL; AF090137; AAD19878.1; -;
DR EMBL; AY339617; AAP88041.1; -;
DR EMBL; BC022567; AAH22567.1; -;
DR EMBL; BC028235; AAH28235.1; -;
DR GO; GO:0016021; C: integral to membrane; TAS.
DR GO; GO:0008146; F: sulfotransferase activity; TAS.
DR GO; GO:0005976; P: polysaccharide metabolism; TAS.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
KW Transferase.
SQ SEQUENCE 411 AA; 46714 MW; F07DOA23B6338A09 CRC64;

Query Match 2.3%; Score 147; DB 2; Length 411;
Best Local Similarity 18.2%; Pred. No. 0.059;
Matches 89; Conservative 58; Mismatches 149; Indels 194; Gaps 16;

Qy 813 WILVIALWFIEL-----LDVWSTC-----SOPICAKWTRTEAGSKSLSS 855
Db 5 WKAVLLALASIAIQYTAIRFTAKSFHTCPLAELAEGLRLC-----EESPTFAYNLSR 59

Qy 856 EGHMDLPDVVITSLPGSAILKQLFPNSSDFEYIRVPTAYIDIPETELEIDSFDVACE 915
Db 60 KTH-----ILITATRSGLSFVGLQFHLNLDVFLPEPL----- 93

Qy 916 WKVSDIRSGHFLRLGWLQSLVQDTKLHLNHLHFNPRGKLAQYFAMNKDKKFKERE 975
Db 94 -----YHVQNTLIPRFTQG-----KSPADRRV 115

Qy 976 SLPEORSOMKGAFFORDAEYI-----RALRRHLVYPSARPV-----LS 1013
Db 116 MGLASRLRLSLYDCDLFLFENYKPPVNNHTDRIFRGASRLVCSRPVCPDPGPADLV 175

Qy 1014 LSSGSWTLLKHLFFQFVLGA-----SMRLYIYVD 1042
Db 176 LEEGDCVRKGLNLTVAEAACRERSHVAIKTVRVPEVNDLRALVEDPRLNLKVIQLVRD 235

Qy 1043 PRMIYSMLYNSKPSLYSLKNVPEHLAKLFKEGKGKCNLNSGVAFYEPFLKELSKS 1102
Db 236 PRGILASRSETFRDITY-----RLWLWYGTGR-----KPNLDVDTQL-TTVCDF 279

Qy 1103 SNAVS--LLSHLWLTAAALRINTDLLPTSQVLKPFEDIVHFPQKTKTERIFAFGLIPL- 1159

Db 280 SNSVSTGLMRPPLKLG-----KYLWLVFEDLARNPMKKTETIYFGLGIPLD 325
Qy 1160 -----SPASINQILFATSTNLFYLYPEGEISPTNTNWKONLPRDEIKLIENI 1207
Db 326 SHVARWIONTRCDPTLKGHKYGTVERN-----SAATAEKWRPRLSYDIVAFQA 375
Qy 1208 CWTLMDRLG 1217
Db 376 CQOVLAQLGY 385

RESULT 15
ID 085160 PRELIMINARY; PRT; 2504 AA.
AC 085160;
DT 01-NOV-1998 (T-EMBLrel. 08, Created)
DT 01-NOV-1998 (T-EMBLrel. 08, Last sequence update)
DT 03-JUL-2004 (T-EMBLrel. 27, Last annotation update)
DE Insecticidal toxin complex protein TcBa (Toxin complex protein).
GN Name=tCBA;
OS Photorhabdus luminescens (Xenorhabdus luminescens).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Photorhabdus.
OC NCBI_TaxID=29488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=W-14;
RX MEDLINE=98306208; PubMed=9641921; DOI=10.1126/science.280.5372.2129;
RA Bowen D., Rocheleau T.A., Blackburn M., Andreev O., Golubeva E.,
RA Bhartiya R., ffrench-Constant R.H.;
RT "Insecticidal toxins from the bacterium Photorhabdus luminescens";
RL Science 280:2129-2132 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=W-14;
RX MEDLINE=21185117; PubMed=11286884; DOI=10.1016/S0966-842X(01)01978-3;
RA Waterfield N.R., Bowen D.J., Fetherston J.D., Perry R.D.,
RA ffrench-Constant R.H.;
RT "The tc genes of Photorhabdus: a growing family";
RL Trends Microbiol. 9:185-191 (2001).
DR EMBL; AF047457; AAC38627.1; -;
DR EMBL; AF346498; AAL18460.1; -;
DR InterPro; IPR002345; Lipocalin.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN 1.
SQ SEQUENCE 2504 AA; 280654 MW; 8P29B1693D1047CE CRC64;

Query Match 2.2%; Score 145; DB 2; Length 2504;
Best Local Similarity 18.1%; Pred. No. 1.2;
Matches 231; Conservative 180; Mismatches 443; Indels 422; Gaps 60;

Qy 29 FSTFEESVSNYSEWA-----VFTDDIDQFKTKQVDFRPNOKLKSMHPSLYFD 78
Db 1003 FTDERYNKRYSRWAGVSELVYPENYVDPTQIRIGQTKMD-ALLQSIQSLNADTVED 1061

Qy 79 AGEIQAMRQKSRASHLHLFRAIRSAVTV-----MLSNPTTYLPPPKHA-----D 122
Db 1062 AFKTY-ITSEQVANKVISAHDNVNVDQLTYFIGIDQAAPGTYYWRSVDHSCENGK 1120

Qy 123 FAA-----KWNIEYGNLPPALYCLLCPEDKVAEFVLEVMYDMRVGYKDWLVENAPGDEV 178
Db 1121 FAANAWGEWNKI-----TCVNPWKNIIIRPV--YMSRL--YLLWLEQOSKKS-D- 1165

Qy 179 PIGHSLTGFATAFDFLYNLLDNHRRQKLEKIM-----VITEEMYEYSKYRSMGKQLLH 232
Db 1166 -----DGKTTIYQ--YNLKLHRIH---YDGSWNTPTFDVTEKVKNTYTSSTDAESL-- 1212

Qy 233 NHOATNMIALITGALVTGDKSKANIWKQAVVDVMEKTMFLLNHIVDGSLEDEGVAGSY 292
Db 1213 -----GLYCTGYQG-----EDTLLVMFYMQSS-----YSSY 1239
Qy 293 T---AKSVTQVFLAQHFNNLD--NNWLKMHFWFYATLLPGFORTVGIADSNYNWF 347
Db 1240 TDNNAPVTGLYIFADMSSDNTNAQATNYWNSY-----PQEDTMADPDS-----1286
Qy 348 YGPESQLVFLDKFLKNGAGNWLAAQIKRHPKDGPMVPSTAQRWSTLHTEYIWDPOLT 407
Db 1287 -----NKKVITRRVNNRYAEDYE-----IPSSV-----TSNSNYSWGDHSLT 1323
Qy 408 POPPADYGTAKIHFFPNWGVVTVYGAGLPN-----TOTNTFVS-FKSGKLGRAV 455
Db 1324 -----MLYGSVPNITFBSAAEDLRLSTNMALSIHNGYAGTRRI 1363
Qy 456 YDIVHFPYSWIDGWSFNGHGHDPQNSFTFAPNGQVVFSEALYGPKLSHLNNVLVPAP 515
Db 1364 QCNLMKQYASLGDKFIITYDSFD--DANRFNLVP--LF---KFGKDENSDDSI CIYNE 1414
Qy 516 SPSSQCNKPHWEGQGECAQMLKWTGE---EVGDAAGEIITASQHGEMVFSGEAVSAYS 571
Db 1415 NPSSE-DKRWYFSSKODNTADYNGGTCIDAGTSNKODFYNNLOEIEVISVTCGYWSSYK 1473
Qy 572 SAMRLKSVYRALLILNLSQTLVVDHIERQEDSPINSVSAFFHNLDIDFKYIPKPMNRYN 631
Db 1474 IS-----NPINI-----NTGIDSAKV--KVTVKAG 1496
Qy 632 GAMMDVMDAHYKMFWDHGHGNSPMASIOEABEKKRWTFQNVVTFOMEST--ITRIA 689
Db 1497 G-----DDQIFTADNSTYVPQ-----QPAPSPFEEMIYQFNLLTIDCKNLNFIDNOA 1542
Qy 690 YVFGPYINVSRCFFIDSSNPGLOISLVNNTTEHVSIIVTDYHNL-----KTRFNVL 741
Db 1543 HIEIDFTATQDGRFLGAETFIIPVTCKVLGTENVIALYSENNGVQVMQIGAYRTRLNTL 1602
Qy 742 GFGGFASVADQG-----QITRFGLTQAIKVPVRHDIRIIFPFGFENIAVGLI 789
Db 1603 FAQQLVSRANGIDAVLSMETQNIQEPQLGAGTYVQLVLDKYDESHGINKSPAIHYVDI 1662
Qy 790 LCISLVILTQWRFYLSFRKLMRWILILVIALWFIEL-----DVMSTCSQPICAKW-----841
Db 1663 FKENDSFVIQGELSETSQTWK-----VFLSYFIEATGNKNHLW-----VRAKYQKET 1711
Qy 842 -----TRTEAE-----CSKKSLSSEGHMDLPDVVITSLPGSCA---875
Db 1712 TDKILFORTDEKOPHGWFLSDHDKTFSGLSSAQAALKNDSEPMDF-----SGANAL 1761
Qy 876 -----ELKOLFNSD--FLYIRVPTAYIDIPELEIDSFVDACEWKV 918
Db 1762 YFWELFYITPMMAHRLLOQNFDAANHFWRYVWNSPGYI-----VDGKIATYHNV 1813
Qy 919 SDIRSGHFRLLRGWLQSLVQDTKLHLQNIHLHEPN-----RGKLAQYFA-----MN 964
Db 1814 -----RPLEEDTSWNAQQLDSTDPDAVAQDDPMHYKVATFMATLDDLMLAR 1858
Qy 965 KDKRKFRRRESLPEQRSQMGAFDRDAEYIRALR-----RHLVYPSARPVLSSSGS 1018
Db 1859 GDAAYRQLERDITLAEAKW-----YTOALNLGDEPQVMLSTTWANPTLGNAASK 1908
Qy 1019 WTLKLHFFQVGLGASMEALYIVRPAWIYSMLYNSKPSLYSLKNVPEHLAKLFKIEGK 1078
Db 1909 TQOQVR--QOVL--TQLRNSRVKTP-----LLGTANSLTAL--FLPQENSKL-----K 1951
Qy 1079 G-----KCNLNSGYAFEPFLRKELSKSNVALLSHLWLANTAALRINTOLLPT 1130
Db 1952 GYWRTLAORMFNLRHNLSDQPLSLPYAKPADPKALLSAAVSAQOGA-----2001
Qy 1131 SYQLVKFEEDIVH-FPQ 1145
Db 2002 --DLPKAPLTIHRFPQ 2015

Search completed: June 23, 2005, 08:49:05
Job time : 138.102 secs